Catastrophic Event Phenomena in Communication Networks: A Survey

by
Christopher Dabrowski
Computer Scientist
Mailstop 8920
National Institute of Standards and Technology
Gaithersburg, MD 20899
(301) 975-3249
christopher.dabrowski@nist.gov

Abstract

With the rise of the Internet, there has been increased interest in the use of computer models to study the dynamics of communication networks. An important aspect of this trend has been the study of dramatic, but relatively infrequent, events that result in abrupt and often catastrophic changes in network state. In the research literature, such catastrophic events have been commonly referred to as *phase transitions*. As interest in phase transitions in communication networks has grown, different approaches to the study of such phenomena have arisen. These approaches are based on differing goals of the researchers, differing investigative methods, and selection of different causal agents to study. While researchers using various approaches have made progress in understanding phase transition phenomena in communication networks, today there is only an incomplete understanding of the overall state of knowledge on this topic and no agreement on a common explanation of how such events occur in communication networks. To provide better understanding of the work done so far, this paper surveys research on phase transitions in communication networks and summarizes what has been learned. The paper identifies four different approaches taken by researchers studying this topic, describes the scope of the work done, identifies the contributions that have thus far been made, and characterizes differences in views on the nature of phase transitions in communication networks. An assessment is also made of weaknesses in the work that has been done, most notably the lack of realism in network models used to date. This survey discusses characteristics of real-world communication networks that need to be included in such models to improve their realism.

Keywords: Internet, Communication Network, Phase Transition, Percolation, Graph Theory.

Contents

1. Introduction

With the rise of the Internet[1], there has been increased interest in the use of computer models to study the dynamics of communication networks. An important aspect of this trend has been the study of catastrophic events that result in an abrupt change in the macroscopic state of an entire network or in a distinguishable sub-network of significant scope. Of most impact are changes in which the network goes from a state in which it is operating normally and communications flow freely to a state where the network is severely degraded or effectively ceases to operate. Such catastrophic events have been commonly referred to in the literature as *phase transitions* from a global operational state to a failed state (Solé and Valverde, 2001; Echenique, Gomez-Gardenes, and Moreno, 2005: Wu, Wang and Yeung, 2008; de Martino et al., 2009; Sarkar et al., 2009). Events of this kind often can occur suddenly, providing no apparent warning before the rapid onset of a change that quickly permeates an entire network and alters its global state. In other cases, the events occur more gradually, suggesting the possibility that they can be predicted. These, and similar, events have been linked to different causes, including excessive load (Solé and Valverde, 2001; Woolf et al., 2002; Arrowsmith et al., 2004; Echenique, Gomez-Gardenes and Moreno, 2005; Lawniczak et al., 2007; Wu, Wang and Yeung, 2008; de Martino et al., 2009; Wang et al., 2009a; and Wang et al., 2009b), propagation of computer viruses (Pastor-Satorras and Vespignani, 2001a; Moreno, Pastor-Satorras and Vespignani, 2002; and Wang et al., 2003; Zou, Towsley, and Gong, 2007), and cascades caused by targeted attacks or failures (Motter and Lai, 2002; Zhao, Park, and Lai, 2004; Watts, 2002; Moreno et al., 2003; Lai, Motter, and Nishikawa, 2004). Despite the potential of such unexpected events to cause widespread economic disruption, the occurrence of phase transitions in real-world communication networks is at best incompletely understood and methods for their prediction are unknown. By communication networks (real-world communication networks), this study refers to the Internet and the world-wide web (www), and significant subsets of these. The study excludes other types of networks (biological, social, voting, etc.) although references by some researchers may occasionally be made to these.

To study global phase transitions in distributed communication systems, researchers have thus far relied on computer models, since use of operational systems to stage disastrous events would be undesirable for obvious reasons, while the cost of developing large-scale testbeds is very high (USC ISI, 2011). The approach taken to the study of phase transitions in computer models has not been uniform, being differentiated by such factors as the disciplinary background of the scientists, by their chosen method of investigation, their research objectives, and by focus on specific causal mechanisms of interest. While the work to date has led to significant findings, it has also been accompanied by shortcomings, and today there is a lack of consensus on an underlying theory that explains how phase transitions occur in communication networks. This paper provides a survey of research on events characterized as phase transitions in communication networks and summarizes what has been learned. The paper discusses different approaches taken by researchers studying this topic, describes the scope of their work, identifies the contributions that have thus far been made by researchers using each approach, and compares and contrasts the different approaches with respect to their views of how phase transitions occur in communication networks. In addition, the survey assesses the weaknesses in the work done so far.

[1] Certain commercial products or company names are identified in this report to describe our study adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products or names identified are necessarily the best available for the purpose.

One important shortcoming of the work done by researchers using all approaches is the lack of realism in the models used to study phase transitions. To overcome this shortcoming, this survey discusses characteristics of real-world communication networks that need to be included in models. The most important of these characteristics are network topologies that capture the AS structure of the Internet and congestion control and routing procedures that are based on real-world protocols. The paper discusses future work to develop realistic models that can be used to more precisely characterize the nature of phase transitions in real-world communication networks. The paper also discusses future work needed to better characterize observed phenomena, such as self-similarity and long-range dependence, in order to understand how these phenomena relate to phase transitions. Better characterization of phase transitions and phenomena that accompany them provides a basis for arriving at a common theory of how phase transitions occur in real-world communication networks. Having a verifiable theory of how global phase transitions occur that also explains observable precursor phenomena in turn provides a basis for developing metrics that can be used to predict the onset of phase transitions. Using such metrics, the widespread effects of undesirable phase transitions can be anticipated in hopes of avoiding the catastrophic impacts that they can have on a communication network.

1.1 Different Approaches to the Study of Phase Transitions in Communication networks

The survey is organized on the basis of four distinguishable approaches taken by researchers to the study of global phase transitions in communication networks. Each approach comprises a distinguishable community of researchers, though there is overlap among them. In all four approaches, the goal of researchers has been to describe the conditions under which phase transitions occur in communication networks, measure the magnitude of these events, and characterize their practical consequences. The four approaches (see Table 1 and Sections 3-6) can perhaps be best distinguished on the basis of two criteria: (1) the choice of which agent or property to study, whose spread through the network leads to the phase transition, and (2) the choice of methods used by researchers to study phase transitions. Examples of causal agents are computer viruses (Pastor-Satorras and Vespignani, 2001a; Moreno, Pastor-Satorras and Vespignani, 2002; Zou, Towsley, and Gong, 2007), events that cause site failure, which propagate through a network (Motter and Lai, 2002; Moreno, Gomez, and Pacheco, 2002; Watts, 2002; Zhao, Park, and Lai, 2004; Lee et al., 2005), and increase in network-wide load and its congestive effects (Solé and Valverde, 2001; Arrowsmith et al., 2004; Echenique, Gomez-Gardenes and Moreno, 2005; Mukherjee and Manna, 2005). The method of study varies by whether it is (a) primarily analytical, proceeding from a specific theoretical framework or employing mathematical analysis techniques to make quantitative characterizations of phase transition behavior (Cohen et al., 2001; Pastor-Satorras and Vespignani, 2001a; Watts, 2002), or (b) primarily empirical, being more strongly based on observations of a computer network simulation which, in many cases, are combined with, or supplemented by, analytical techniques or examination of real-world data, from which conclusions are drawn (Motter and Lai, 2002; Echenique, Gomez-Gardenes and Moreno, 2005; Mukherjee and Manna, 2005; and Lee et al., 2005). Though many researchers use both analytical and empirical methods; most rely primarily on one and use the other on a supporting basis.

The first of the four approaches (described in Section 3) considers the problem from the standpoint of percolation theory of random graphs. The primary goal of the percolation-based approach is to extend percolation models, supplemented with mean-field equations, to develop analytic estimates of the threshold at which phase transitions occur, and the magnitude of the change. The models used in the percolation-based approach are highly abstract and the causal agent is usually unspecified. The second approach (Section 4) is the epidemiologically-based approach, which combines percolation theory and epidemiologic modeling of disease spread. The goals of the epidemiologically-based approach are largely the same, though there is also an emphasis on understanding, and controlling, the effects of disease-like spreading agents, which lead to phase transitions. The third approach (Section 5) studies the dynamics of cascades (also referred to as avalanches) that lead to phase transitions. The emphasis of these studies is on developing a detailed representation of the cascading mechanism itself and in understanding how it spreads. Some researchers in this group use percolation theory as a basis to analytically estimate thresholds, while others use a primarily empirical approach to derive threshold estimates and characterize phase transition behavior on the basis of observations of simulated cascades. To some extent, the epidemiologically-based approach and cascade studies can be viewed as specializations of the percolation-based approach. This is because while each of these approaches employs percolation theory, in contrast to the percolation-based approach, the epidemiologically-based approach and the cascade studies focus on the effects of specific spreading mechanisms. The fourth approach (Section 6), which has resulted in many publications, is based on simulation of communication networks and in some cases, reliance on real-world data. Researchers in this group study phase transitions caused by the growth of congestion in networks. The approach used here is primarily empirical, based on observation of model simulations; however, in a number of cases, analytical means, most notably mean-field theory equations, are also used. In contrast to the percolation-based and epidemiologically-based approaches and also to some extent the cascade studies, researchers who studied global phase transitions caused by congestion did not rely nearly as much on percolation theory. These differences suggest a need for investigations designed to produce a common understanding of how phase transitions occur in communication networks.

Table 1. Summary of four approaches

Focus of Approach	Method of Investigation	Causal Agent
Abstract percolation theory	Primarily analytical, proceeding from percolation theory	Spreading generic (unspecified) property, which fails sites
Epidemiologic spread	Primarily analytical, proceeding from combination of epidemiologic model of disease spread and percolation theory	Spreading computer virus, which fails sites
Studies of cascades or avalanches	Either (1) analytical, proceeding from percolation theory or (2) empirical, based on observation of simulation	Quantitative cascading property which fails sites
Network congestion	Primarily empirical, based on observation of simulation	Excessive load, which jams network flow

1.2 Scope and Definitions

This paper focuses on work that studies phase transitions in wired communication networks and their related Internet topological structures, excluding systems such as wireless networks, which have a different structure and dynamics and are best treated separately.

In all the works described in this survey, communication networks are modeled using graph theory concepts. Therefore, before proceeding, it is desirable to provide selected definitions relating to graph topologies that will be used throughout this paper. For more detailed treatments, see (Newman, 2003; Boccalettia et al., 2006; Dorogovtsev, Goltsev, and Mendes, 2008; da Fontoura Costa et al., 2011). A *graph* $G = (N, L)$ consists of two sets N and L. The elements of $N \equiv \{n_1, n_2, \ldots, n_N\}$ are the *nodes* (or *vertices*) of the graph G, while the elements of $L \equiv \{l_1, l_2, . . . , l_K\}$ are its *links* (or *edges*), where each l_i consists of a pair of elements of N. If a link can be traversed in one direction, the link is considered directed, and the graph is a *directed graph*. If a link can be traversed in both directions, the link is considered bi-directed, and the graph is a *bidirected graph*. The graph also can be represented as having *undirected* links, which can be traversed in either direction (Newman, 2003). In the majority of cases, the Internet is, in practice, undirected—as most surveyed authors indicate explicitly or implicitly. The WWW can be directed, though in (Cooper and Frieize, 2003) an undirected model of the WWW is presented. In this paper, the term *site* will be used in preference to *node*, since the term site more accurately conveys the concept of distinct location in a communication network which contains a combination of hardware and software components, while the term node (or vertex) is more abstract. In addition, the term *network* will be equivalent to the term graph, when these terms are used to describe simulation models. Graphs can be either finite or infinite: where possible, the distinction is noted. Simulations were always conducted in finite networks. Links will be of equal weight, or unweighted, in different works unless otherwise noted. A few studies attach unequal weights to links to denote strength of attachment or distance of the sites across the Internet.

Two sites that are joined by a link are referred to as *adjacent* to each other. A link is *incident* with sites, i and j, if it joins these two sites. In this paper, all graphs are assumed to be traversable in either direction along a link that joins two sites. A *subgraph* $G' = (N', L')$ of $G = (N, L)$ is a graph such that $N' \subseteq N$ and $L' \subseteq L$. The *degree* (or *connectivity*) k_i of a site i is the number of links incident with the site, and is defined in terms of the adjacency matrix[2] A as $k_i = \sum_{j \in N} a_{ij}$, where each a_{ij} is an element of A. Another important concept used in this survey is the *statistical ensemble* for a graph, G, having N sites and $|L|$ links, where a link between two sites exists with a prescribed probability p (and is absent with a probability $1 - p$). In (Newman, 2003), each member of the statistical ensemble for G is defined as a unique graph with $|L|$ links, which is realized with a probability $p^L(1 - p)^{M-L}$, where $M = ½|N|\,(|N| - 1)$ is the maximum number of links[3]. (An ensemble average, denoted by the angled brackets ‹›, is the expected value of some quantity over the members of the ensemble. As an example, $\langle k \rangle$ denotes the expected degree across all members of a statistical ensemble of a graph. These definitions are common to all work surveyed here. Additional definitions will be provided below where necessary. In a few cases, the same symbol will be used to define different quantities, though these instances are widely spaced. This is done to maintain consistency with well-known definitions used in different approaches.

The papers surveyed here concern themselves to a great extent with simulated networks that are based on random graph topologies. There are different types of random graphs, and the exact meaning of the term *random graph* depends on the definition of the particular type. As an example, the earliest random graph model, known as the Erdös-Rényi random graph (Erdös and Rényi, 1961), can be succinctly understood as being a statistical ensemble for a graph $G = (N, L)$, in which each member in the ensemble has an equal probability of realization. Starting with a disconnected graph G, an Erdös-Rényi random graph (i.e., an individual member of the statistical ensemble) is generated by connecting pairs of randomly selected sites from N until the number of links equals $|L|$, (prohibiting duplicate links between

[2] An adjacency matrix A, can be defined as $A = a_{ij}$, where each a_{ij} is a member of A, if $a_{ij} = 1$ when there is a link between sites i and j, and $a_{ij} = 0$ otherwise. When the graph is undirected, A is symmetrical. If A is directed, it may not be symmetrical.

[3] Alternatively, it is possible to define an ensemble for a graph G that consists of all graphs that have exactly $|N|$ sites and $|L|$ links. For example, consider a graph $G = (N, L)$ in which N consists of three sites $\{n_1, n_2, n_3\}$ and L consists of two links which may be configured three ways to join pairs of sites. It is easy to see that there are three ways in which sites n_1, n_2, and n_3 may be joined by two links. Thus, the ensemble has three members, each of which is realizable with the same probability. See also (Dorogovtsev, Goltsev, and Mendes, 2008) for examples.

the same sites). Additional types of random graphs will also be defined in subsequent sections, as needed.

An important way in which different types of random graphs are distinguished is by differences in degree distribution. The fraction of sites in a network that has degree k is defined as p_k, which is also understood to be the probability that a site has the degree k. In a network modeled as an *Erdös-Rényi random graph*, each link is present with the same probability, and the degree distribution is known to have a Poisson distribution (Newman, 2003). In contrast, a scale-free random graph (Barabási and Albert, 1999), or *scale-free network*, is characterized by a degree distribution defined as

$$p_k \sim k^{-\alpha} \tag{1}$$

for some constant exponent α. In contrast to the *Erdös-Rényi* random graph, this distribution is heavily skewed so that a few sites have many links incident upon them, while the vast majority of sites have far fewer links. The topology for such a graph features a few highly connected sites, which are referred to as *hubs*. The plot for the frequency distribution of Equation (1), shown in Figure 1, is said to have the property of being scale-free, or scale-invariant, because it maintains a constant slope across powers, or scales, of k. Often, scale invariance is referred to as *self-similarity* (Solé, 1996; Hinrichsen, 2006), which can be intuitively understood as a pattern, or observed trend, that cannot be statistically distinguished at different scales. Self-similarity in data associated with communication networks is an important phenomenon that has been studied by a number of researchers, as we shall see below.

One particular type of scale-free network, known as the Barabási-Albert scale-free network, is created using a *preferential attachment* growth algorithm (Barabási and Albert, 1999). For $G = (N, L)$ starting with some small set of connected sites, G', where G' is a subset of G, at every succeeding time step, a new unconnected site is selected from a G and connected to site i within G' of the already existing sites. The probability Π_i that the new site is connected to site i depends on the connectivity k_i of site i, such that $\Pi_i = k_i / \sum_{j \in N} k_j$. Thus, there is a higher probability that each new site, upon creation, is connected to an already highly connected site, resulting in a power-law degree distribution.

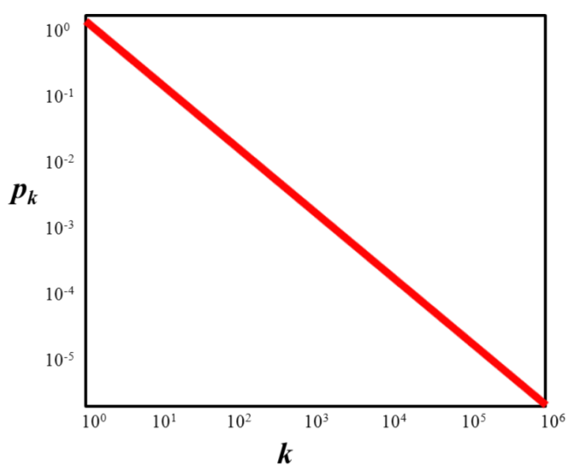


Figure 1. Conceptual representation of a scale-free frequency distribution of degree k, defined by Equation (1). In this example, the value of α is 1, although other values of α are, of course, possible. A frequency distribution is the count of the occurrences of values of k.

Finally, it is necessary to briefly discuss the topic of degree correlation. A graph is correlated with respect to a degree, k, if the probability that a node of degree k is connected to another node of degree, k', depends on k. Following (Boccalettia et al., 2006), degree correlation may be expressed as a *conditional* probability that a site of degree k is connected to another site of degree k', written $P(k'|k)$, where $\sum_{k'} P(k'|k) = 1$, and it is the case that $k\,P(k'|k)\,P(k) = k'\,P(k|k') / P(k')$. If no degree correlation exists, i.e., $P(k'|k)$ does not depend on k, it has been shown that $P(k'|k) = k'\,P(k') / \langle k \rangle$. If a graph is correlated[4] with respect to degree k, it is possible to calculate the average degree of the nearest neighbors of sites with degree k, denoted $k_{nn}(k)$, where $k_{nn}(k) = \sum_{k'} k' P(k'|k)$. If no degree correlation exists, i.e., $k_{nn}(k)$ is independent of k, then $k_{nn}(k) = \langle k^2 \rangle / \langle k \rangle$. Beyond these formulae, additional descriptions of computing degree correlation exist (Mahadevan et al., 2006; Piraveenan, Prokopenko, and Zomaya, 2009). Most work surveyed in this paper assumes the absence of degree correlation, though the phenomenon has been found in real-world networks, such as the Internet (see Section 7).

1.3 Phase Transitions

An early statement of the concept of a phase transition can be found in the study of thermodynamic systems in statistical physics. It is a concept of some prominence in this survey and is common to many areas of study. Therefore the phase transition requires explanation. In the literature on thermodynamic systems, phase transitions are generally understood as changes between *macroscopically different system-wide equilibrium states*. A simple example of a phase transition is water undergoing the change between gaseous, liquid, and solid states. More precisely, a phase transition between states is expressed as a change in the value of a variable known as the *order parameter* (Jaeger, 1998). An example of an order parameter for the state change of water is density, which changes measurably when, for instance, water freezes (becomes solid) or when water evaporates (becomes gaseous). Other order parameters, which can be considered to measure the state of a system (such as water) are pressure, temperature, internal energy, entropy, or number of particles in the system. Systems in which phase transitions have been studied include substances such as water, Helium (Jaeger, 1998), the emergence of order in ferromagnetic fields (Dorogovtsev, Goltsev, and Mendes 2008), among others.

In thermodynamic systems, the free energy of a system is a key element. In such as system, free energy is often expressed in terms of the *Gibbs free energy potential*[5]; where in a simple, two-dimensional random graph, free energy may represent the distribution of clusters of sites, or subgraphs, and their mean size (Nakanishi and Stanley, 1978). (Thus, a phase transition occurs in a graph when this distribution changes radically, or percolates, but more on this later). As originally stated (Ehrenfast, 1933), it is possible to classify phase transitions by the behavior of the first and second derivatives of the Gibbs free energy potential. If some of the first derivatives are discontinuous, the phase transition is said to be discontinuous or *first order*. If all the first derivatives are continuous, but discontinuities appear in

---

[4] In the case where degree correlation is positive, if sites tend to connect to sites with similar degrees, $k_{nn}(k)$ is an increasing function of k, and the graph is considered to be assortative. If sites with low degrees are more likely to connect to sites with high degrees, $k_{nn}(k)$ is a decreasing function of k, and the graph is considered disassortative (Boccalettia et al., 2006).

[5] In (Jaeger, 1998), the Gibbs free energy, Z, is given as $Z(T, p) = G = U - T\,S + pv$ (where T is temperature, S is entropy, p is pressure, v is volume and U is free energy as described).

the second derivatives, then the phase transition is said to be continuous or *second order*[6]. Also of importance is that the behavior of these derivatives is expressed theoretically only in the so called *thermodynamic limit* (ben-Avraham and Havlin, 2000; Dorogovtsev, Goltsev, and Mendes 2008), where the system is theoretically defined to be infinite in size and to have infinite valued variables. Phase transitions themselves occur only in the thermodynamic limit. Therefore, in the observation of finite real-world systems, or simulations, phase transitions can more properly be said to have been observed, or likely to have occurred, but cannot be verified without theoretical derivation.

The theory of phase transitions in thermodynamic systems has been used in the study of phase transitions in communication networks. Researchers in communication networks have applied percolation theory concepts to random graph models (mostly in the percolation-based approach and the epidemiologically-based approach and partially in the cascade studies). As will be discussed further below, many researchers have viewed phase transitions, or percolation transitions, in their network models as continuous, or second-order transitions, which can be shown mathematically to occur in the thermodynamic limit (although in some cases discontinuous, or first-order transitions, were also observed). Researchers who studied catastrophic failures arising from congestion have also characterized the global failures they observed in simulations as appearing to be phase transitions. However, in this case the observations made were primarily empirical in the context of finite systems, and could not be always verified theoretically as occurring in a thermodynamic limit, though work by (Sarker et al., 2009; Rykalova, Levitan, and Brower, 2010; Sarkar et al., 2012) uses the concepts and terminology of phase transitions more extensively for an observed system. As we will see, researchers who studied congestion observed evidence of both continuous and discontinuous phase transitions.

The classification of continuous vs. discontinuous phase transitions is of importance to this study for another reason. In the case of the former, a more gradual change from one state to another occurs in some limited time range, during which the beginning of the change may be preceded by detectable precursor behavior. This precursor behavior, known as *critical slowing down* (Solé et al., 1996), generally manifests itself when a system approaching a phase transition is perturbed from equilibrium, and takes increasingly long to return to equilibrium as the system gets closer to the point of transition. Thus, the presence of critical slowing down is potentially of use in determining whether a phase transition can be predicted or not. Critical slowing down appears to be absent from systems that undergo continuous phase transitions.

---

[6] The classification of phase transition order in thermodynamic systems is discussed in (Jaeger, 1998). Briefly summarized, the original classification in a 1933 article by Paul Ehrenfast (Ehrenfast, 1933) states that when any of the first derivatives of the Gibbs free energy equation for Helium is discontinuous, or for example a jump in the first derivative of entropy, $S = -(\partial G / \partial T)p$, occurs at the transition point, the surface of *Z* is "kinked" and the phase transition is discontinuous or first order. When the first derivatives are continuous, but a second derivative for Helium is discontinuous, such as for $c_p = (\partial^2 G / \partial_T^{\,2})_p$ where c_p, (i.e., *heat capacity*) jumps at the transition point, the transition is considered continuous, or second order. Here, the origin of the concept is attributed to (Ehrenfast, 1933). Jaeger's translation of Ehrenfast's concepts describes "the Ehrenfast classification", which was originally developed to explain phase transition order on the basis of first and second derivatives of the surfaces of the "Gibbs free energy" equation for the transition of Helium from liquid to gas state. Since Ehrenfast's original work, the concept of phase transition order has been generally accepted and further extended and modified, while being applied to other types of thermodynamic systems, particularly magnetism (Kadanoff, 2009) and random graphs (Nakanishi and Stanley, 1978, Newman, 2003). The terminology of phase transition order has been widely used since then, including in the literature on catastrophic events in communication systems.

As will be discussed in Sections 3 and 4, a random graph is also a system which may undergo a phase transition from a state in which most of its sites are disconnected from each other to a state in which most sites belong to a single connected cluster. The process by which this change occurs is also percolation. The order parameter measures extent of global connectivity in the network (i.e., all sites belong to a single cluster). As we shall see, the salient characteristics of phase transitions in thermodynamic systems have been argued to have been observed in the percolation of networks. These characteristics have been observed both in networks in which congestion increases leading to global failed states (i.e., network connectivity disintegrates into many disconnected clusters of sites). These characteristics have also been observed in networks in which network connectivity grows and a single cluster of connected sites is observed to emerge.

1.4 Organization of This Paper

The contents of the paper are organized as follows. Section 2 discusses previous surveys of work on phase transitions in communication networks. Here, related studies are identified, though none covers the ground of this paper. Sections 3 through 6 describe the four approaches to the study of global phase transitions, summarizing the contributions each has made. The disadvantages and weaknesses of each approach are also identified. Section 7 provides an assessment of the overall state of knowledge on phase transitions in communication networks and discusses needed future work. Future directions focus on the real-world elements that need to be included in models of communication networks, in developing a common theory of phase transitions in networks, and on commencing work on metrics for predicting phase transitions. Section 8 concludes. An appendix is provided to further explain self-similarity and long-range dependence.

2. Related Studies

Several previous works survey the literature of phase transitions in communication networks. These previous studies either do not cover the entire range of works surveyed in the present paper or do not address the topics outlined in the introduction. Among these is (Smith, 2011), which conducts a detailed review and discussion of various manifestations of self-organization phenomena in communication networks, including self-similarity, long-range dependence, oscillations, periodicities, and phase transitions. However, (Smith, 2011) provides only a limited review of work on phase transitions in communication networks, most notably omitting the bulk of the work on approaches based on percolation theory and epidemiologic models. The survey provided by the present paper is more complete in this regard, but also focuses on one class of phenomena: network-wide phase transitions.

This survey focuses on the more general topic of known theoretic properties of network structures represented as various types of random graphs, including those properties related to, in part or whole, percolation phase transitions. The brief paper by Barabási and Albert (Barabási and Albert, 1999) omitted the bulk of the work on congestion studies of communication networks. In addition, a great deal of work has occurred in this area since the publication of (Barabási and Albert, 1999). Surveys on complex networks by (Newman, 2003) and (Boccalettia et al., 2006) provide detailed overviews of the properties of random graphs and, to some extent, known statistical properties of real-world networks as well. However, these two works also have a broad scope with respect to complex phenomena in networks. Though three of the four approaches discussed above (the first three in part or whole based on percolation) are identified in (Newman, 2003), and all four are actually mentioned in (Boccalettia et al., 2006), the discussion of phase transitions in communication networks in both is limited to review of a small subset of papers in each category. A more recent paper by (Dorogovtsev, Goltsev, and Mendes, 2008) also surveys work on statistical mechanics of complex network, focusing on description of complex phenomena, most notably percolation transitions and non-equilibrium phase transitions. While this work provides a comprehensive overview of the theoretic results related to phase transitions, empirical studies of communication networks using simulation models are also very much under-represented, as in (Barabási and Albert, 1999). A useful paper by (da Fontoura Costa et al., 2011) surveys work on simulation and computer modeling in 22 different types of real-world networks, including communication networks and the Internet. However, this paper does not focus on the issue of phase transitions, and so the topic is not treated in depth.

The present survey is specifically directed at the phenomenon of phase transitions and treats all approaches to the study on this topic in greater depth than the previous surveys. This survey provides a detailed treatment of the findings about phase transitions in communication networks, which were made by researchers using each of the approaches. In particular, this survey identifies important defined quantities that characterize phase transitions and network-wide phenomena, which have been identified by the four approaches. These important quantities and phenomena are summarized in Table 2.1 and Table 2.2, respectively and will be elaborated in ensuing sections. In addition, this survey provides an assessment of the shortcomings of each approach. On the other hand, previous surveys have been limited to covering a subset of the approaches identified here or to discussion of an overview of each approach and include phenomena besides phase transitions. Also in contrast to previous surveys, this survey discusses in detail the prospects for studying phase transitions using models with characteristics of real-world networks. This survey identifies work that describes real-world characteristics, which are relevant to developing improved models of communication networks in which global phase transitions can be studied.

Table 2.1. Important quantities related to phase transitions which were studied by researchers who used the four approaches discussed in this survey. The term giant connected component is abbreviated GCC.

Important Measured Quantities	Approach			
	Percolation	Epidemiologic	Cascade	Congestion
Spreading property that occupies sites	Occupation of sites by generic property and emergence or disintegration of GCC	Occupation of sites by virus or disease agent and emergence or disintegration of GCC.	Occupation of sites by cascading agent and emergence or disintegration of GCC.	Sharp decrease in system throughput and sharp increases in packet lifetime (duration for packet transmission), number of packets in the system, and queue lengths at sites.
Estimation of critical point for phase transition	Measured	Measured	Measured	Measured
Growth rate of GCC	Measured	Measured	Measured	Not measured
Distribution of non-GCC components	Measured	Not measured	Not measured	Not measured
Distribution of inter-site distances	Measured	Not measured	Not measured	Not measured

Table 2.2. Network-wide phenomena associated with phase transitions which were studied by researchers who used the four approaches discussed in this survey. The term giant connected component is abbreviated GCC.

Phenomena	Approach			
	Percolation	Epidemiologic	Cascade	Congestion
Critical slowing down				In (Sarkar et al., 2012)
Phase transition order	Continuous in emergence and disintegration of GCC.	Continuous in emergence of GCC	Continuous and discontinuous in formation of GCC. Continuous and discontinuous in disintegration of GCC or largest component.	Continuous and discontinuous phase transition from free to congested state.
Self similarity in measured quantities	In GCC growth above criticality, distribution of non-GCC components at criticality, distances between sites at criticality.	In GCC growth above criticality.		In packet queue lengths at sites, packet lifetimes, packet delay times, distribution of congestion durations, fluctuation in number of packets at sites at criticality and selected points below criticality.
Long-range dependence in measured quantities				In packet queue lengths at sites, packet delay times at criticality and selected points below criticality.

3. Modeling Catastrophic Events in Communication Networks using Percolation Theory

In this approach, catastrophic events in communication networks are studied by using abstract percolation theory to represent phase transitions in various types of random graphs. Originally adapted from the study of lattice structures, percolation is the process by which some property of interest spreads through a graph. Different kinds of percolation processes have been studied in lattices (ben-Avraham and Havlin, 2000) and random graph structures (Erdӧs and Rényi, 1961; Bollobás, 1984; Molloy and Reed, 1995; Cohen et al., 2000; Cohen et al., 2001; Newman, Strogatz, and Watts, 2001; Cohen, ben-Avraham, and Havlin, 2002; Bollobás, Janson, and Riordan, 2006; and Dorogovtsev, Goltsev, and Mendes, 2008). Percolation has also been studied from the epidemiological point of view (Pastor-Satorras and Vespignani, 2001a among others, see Section 4) and the point of view of cascades (Watts, 2002 among others, see Section 5)—and contributions made from both. The concepts of percolation are universal to many scientific fields and areas of study.

Notable percolation concepts include site percolation, in which the spreading property progressively occupies sites within a network, and bond percolation, which occupies links (ben-Avraham and Havlin, 2000; Moore and Newman, 2000). In this survey, only site percolation is considered in detail, since this is the predominant form of percolation considered by researchers who studied phase transitions in random graph representations of communication networks. The application of percolation theory to communication networks was motivated by an interest in the reliability of the Internet and its resilience to attack (Albert, Jeong, and Barabási, 2000; ben-Avraham and Havlin, 2000; Cohen et al., 2000; Cohen et al., 2001; Cohen, ben-Avraham, and Havlin, 2002; Callaway et al., 2000; Newman, 2003). Therefore, researchers covered in this section, and the ensuing sections, defined variables and equations that can perhaps be used as measures of observed percolation in real networks. Generally, these researchers modeled the Internet using random graphs. Originally described by Erdӧs and Rényi (Erdӧs and Rényi, 1961), the percolation of random graphs is generally understood to be a thermodynamic process, of the kind discussed in Section 1.3 (ben-Avraham and Havlin, 2000; Dorogovtsev, Goltsev, and Mendes, 2008; Newman, 2003). Definitions and descriptions of percolation are also provided in (Grimmett, 1989; Stauffer and Aharony, 1994; ben-Avraham and Havlin, 2000; Bollobás and Riordan, 2006).

3.1 Basic Concepts

In site percolation, a site possesses a spreading property of interest with a probability, p. If p increases so that it exceeds some critical probability, p_c, known as the *percolation threshold* or *critical point*, i.e., $p > p_c$, a *percolation transition* occurs. When such a transition occurs, a *giant connected component* emerges that consists of connected sites having the property of interest. The proportion of network sites within the giant connected component is represented by P_∞. If p continues to increase above p_c, P_∞ also increases. Below p_c, the graph is composed many smaller, isolated components, and no giant connected component exists (i.e., the value of P_∞ is 0 for $p < p_c$). At p_c, the giant connected component comes into existence and $P_\infty > 0$. The giant component is a unique component that emerges at p_c and continues to grow as more sites in the network are occupied by the critical property (i.e., $P_\infty > 0$ at p_c). Thus P_∞ serves as an *order parameter*, which characterizes the state of the network above p_c. If the growth of the giant component is unrestricted, all other connected components in the network will decrease in size and number, and ultimately, the giant component will connect the entire network. Figure 2 shows the concept. As in other thermodynamic systems, the percolation transition in a random graph takes place in the thermodynamic limit, in which the network is assumed to have infinite size and the giant connected component itself is infinite, as indicated by the symbol, P_∞.

As an example of a percolation transition in a finite communication network, Figure 2 might represent the booting up of a network, in which individual sites become operational and connect to their neighbors. Thus, p represents the probability that a site is operational, i.e., it is capable of sending and receiving data. As p increases, small connected components of communicating sites begin to form. When $p > p_c$, the giant connected component forms, which consists of operational sites that communicate with each other. As p continues to increase, the giant component continues to grow, and the operational network also grows larger. Hence, the presence of the giant component signifies that the network itself is in an operating state.

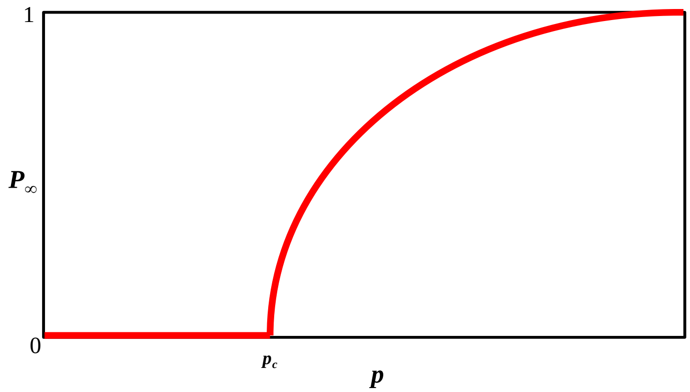


Figure 2. A conceptual representation of the percolation transition is shown, in which p, the proportion of sites occupied by the spreading property is plotted on the horizontal axis, while the probability P_∞ that a site belongs to the giant connected component, is plotted on the vertical axis. P_∞ undergoes a phase transition at the critical probability threshold, p_c. Below p_c, $P_\infty = 0$, and the giant connected component does not exist. Above p_c, P_∞ grows at the rate indicated by Equation (2). The figure was adapted from (ben-Avraham and Havlin, 2000).

For $p > p_c$, the growth rate of the giant connected component, P_∞, is given by:

$$P_\infty \sim (p - p_c)^\beta. \tag{2}$$

Equation (2) is a known power-law relationship (ben-Avraham and Havlin, 2000), which possesses the property of scale-invariance or self-similarity, discussed in Section 1.1. In this equation, β is referred to as a *critical exponent*, whose value varies on the basis of the space dimensionality, or connectivity, properties of a graph. In addition to Equation (2), other power-law relationships are known to exist for other quantities at or near the critical point for a percolation transition. For instance, the distribution of cluster sizes at the critical point is known to behave as:

$$n_s \sim s^{-\tau}, \tag{3}$$

where n_s is the number of clusters of size s as $s \rightarrow \infty$, and τ is also a critical exponent whose value varies on the basis of the dimensionality properties of the graph. Away from the critical point, the power-law relationship given in Equation (3) no longer holds. Still another power-law relationship exists for the correlation length, ξ, which below the critical point, approximates the typical distance between sites in a cluster[7]. At and above the critical point, ξ diverges as given by:

$$\xi \sim |\, p - p_c|^{-\gamma}, \tag{4}$$

[7] The correlation length also has been defined as the upper cutoff of the radius of those clusters which contribute to the mean cluster size near the percolation threshold (Christensen, 2002).

where γ again depends on dimensionality properties. In an infinite-sized network, for distances larger than ξ, distances between sites in the giant connected component no longer follow the power law, and the component is said to become homogenous. Further discussion of power-law relationships for percolation transitions can be found in (Stauffer and Aharony, 1994). The appearance, and disappearance, of scale invariance or self-similarity in various measured quantities as a network approaches, and reaches, the critical point is in the study of phase transitions in networks, and will be discussed further below. As will be shown, scale invariance was studied by researchers using other approaches for a variety of quantities listed in Tables 2.1 and 2.2, in addition to those in Equations (2–4).

Finally the percolation transition in a network is very often a continuous phase transition, (ben-Avraham and Havlin, 2000; Newman, 2003; Dorogovtsev, Goltsev, and Mendes, 2008; Buldyrev et al., 2010), though there have been exceptions to this conclusion, as discussed further below. In particular, some authors have reported circumstances in which discontinuous phase transitions occur (Watts, 2002; Buldyrev et al., 2010; Li et al., 2012) and others.

3.2 Using Percolation Theory Concepts to Study Reliability of Communication Networks

Percolation theory, though conceived in the context of an infinite network, can be used to approximate many different processes in finite communication networks, besides the spread of connectivity in the previous section. For instance, in addition to the establishment of widespread network connectivity, the observed percolation transition can be used to represent the spread of viruses and cascades (see Sections 4 and 5, respectively). Of great interest to researchers has been the study of the breakup of the giant connected component, which serves as a model for the destruction of global network connectivity. Consider events such as failures or attacks that cause a site to become inoperable. In this case in a random graph model, such events can be represented by the removal from the graph of the affected sites, so that the network becomes less operational, and p, decreases. If p falls below p_c, the giant connected component undergoes the reverse of the phase transition described above and disintegrates. The disappearance of the giant connected component of operable sites then represents a phase transition of the communication network to a global inoperable state. This event, and the value of p_c at which it occurred, was of special interest for the study of network reliability. Many researchers who used percolation theory analyzed network models that were based on well-known random graph topologies and verified their findings through simulation (Barabási and Albert, 1999; Cohen et al., 2000; Cohen et al., 2001; Cohen, ben-Avraham, and Havlin, 2002; and others). Much of their work involved deriving expressions for estimating important quantities related to the phase transitions and verifying the existence of power-law relationships given by Equations (2–4) as summarized in Tables 2.1 and 2.2. Two quantities of interest are covered here briefly: the critical percolation threshold, p_c, and the proportion of networks sites in the giant connected component, P_∞.

3.2.1 Estimating the Critical Percolation Threshold

The estimation of p_c, for example, was studied by (Albert, Jeong, and Barabási, 2000; Cohen et al., 2000; Callaway et al., 2000) for a randomly connected graph representations of a communication network, focusing on graphs with inhomogeneous degree distributions. In (Cohen et al., 2000) for a randomly connected graph, each site has a connectivity of k (i.e., the number of connections emanating from a site). Here, it was found that $1 - p_c = 1 / (\kappa_0 - 1)$, where $\kappa_0 = \langle k_0^2 \rangle / \langle k_0 \rangle$ and $\langle k_0 \rangle$ is the ensemble average value, computed from the initial connectivity distribution prior to the beginning of the disintegration of the giant connected component, discounting links that formed cycles. In practical terms, this meant that if each site in the giant component was connected to at least two other sites (excluding cycles), the component would remain intact, a result which suggested a fair degree of robustness for networks based on random Erdös-Rényi random graphs with a Poisson connectivity distribution. This condition was found to hold for a finite random graph, when the average ensemble degree for the graph, $\langle k \rangle$, was equal to 1 (Cohen et al., 2000).

Of special interest were scale-free networks, in which connectivity followed a power-law distribution, $P(k) = ck^{-\alpha}$. This class of graphs was widely believed to represent the topology of the Internet, which was thought to be composed of relatively few, highly connected hubs, where $\alpha \approx 2.5$. In (Cohen et al., 2000) it was shown that for random untargeted attack on a finite scale-free network with $\alpha < 3$, a giant component remained in place even if nearly 100 % of the sites were removed (although it would be comprised of a very small proportion of the original population of sites). If a scale-free graph theoretically approached infinite size with respect to the number of sites and links, it was shown that $p_c \rightarrow 0$: in other words, the giant connected component was always guaranteed to exist. In practical terms for a finite scale-free network, the threshold, p_c, for the existence of the giant connected component was found, when the network was very small, i.e., the network would be operational even if very few sites were functioning[8]. This situation is illustrated in Figure 3 for a finite scale-free network, where the variable q represents the inverse of p, or $q = 1 - p$.

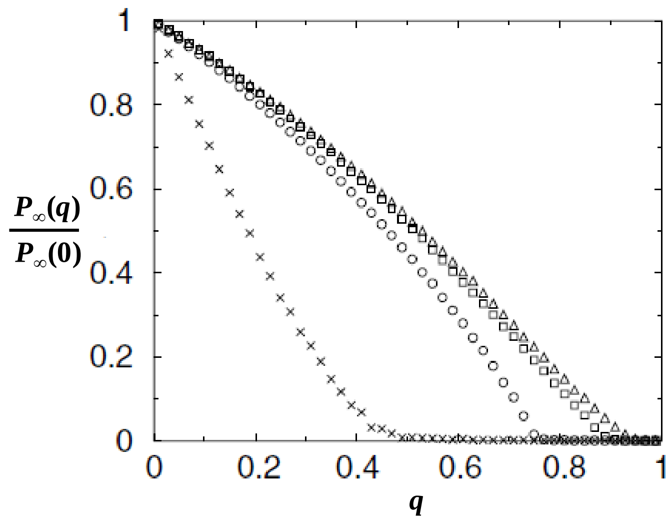


Figure 3. Phase transition to failed state for networks having scale-free connectivity distribution. The graph shows the fraction of sites that remain in the giant connected component after breakdown of a proportion q, $q = (1-p)$ of sites, where the fraction of remaining sites is represented as a function $P_\infty(q) / P_\infty(0)$. Separate curves show the breakdown for $\alpha = 3.5$ (crosses) and $\alpha = 2.5$ (other symbols), as obtained from computer simulations of a network having up to 10^6 sites. For $\alpha = 3.5$ and $q \approx 0.5$, the giant component disintegrates, and the network becomes fragmented. However, for $\alpha \approx 2.5$ (believed to be the degree distribution of the Internet), the giant component continues to exist at higher values of q. In this chart, the three curves for $\alpha \approx 2.5$ represent simulations in which, K, the maximum degree of any site in the network, i.e., the most connected hub site is varied. By increasing K, the growth in the size of the network is approximated. Different curves for $K = 25$ (circles), $K = 100$ (squares), and $K = 400$ (triangles) illustrate the effect of increasing K on q, and on the percolation threshold, $p_c = 1 - q$. The rightward shift in q, $q \rightarrow 1$, illustrates that the critical threshold, p_c, decreases as K grows, which supports the theoretical prediction that $p_c \rightarrow 0$ as network size becomes infinite. The figure is from (Cohen et al., 2000).

[8] However, the proportion of sites which are within the giant connected component of a scale-free network, i.e., the size of the operational network, would be relatively small, even though p is above a very small p_c.

This analysis led to the conclusion that the connectivity of the Internet was highly resilient to catastrophic events in the form of phase transitions to failed states which were caused by random attack on sites. Analyses showed similar results for random failure of nodes in a scale-free network (Albert, Jeong, and Barabási, 2000; Callaway et al., 2000). However, in (Cohen et al., 2001) as well as in (Albert, Jeong, and Barabási, 2000; Callaway et al., 2000), it was found that in the case of intentional attacks aimed at hubs of finite scale-free graphs where $\alpha > 2$, even attacks that eliminated only a few hubs would cause the disintegration of the giant connected component and a global phase transition to a failed state. Hence, it was concluded that the Internet was vulnerable to targeted attack on hubs.

3.2.2 Estimating the Size of the Giant Connected Component in Scale-Free Graphs and Other Variables

To further define variables and equations that can be used as measures of observed percolation, researchers also estimated the size of giant connected component in scale-free network models at, and above, the percolation threshold. They also estimated the values of critical exponents for power-law relationships in networks modeled as graphs. Estimates of the proportion of sites in the giant connected component, P_∞, in networks modeled as random graphs were provided in (Erdös and Rényi, 1961; Molloy and Reed, 1995; Callaway et al., 2000). For a scale-free network ($\alpha > 3$) with a degree distribution, $P(k) = ck^{-\alpha}$, the proportion of sites in the giant component, P_∞, for $p > p_c$ was approximated in (Cohen, ben-Avraham, and Havlin, 2002) to be $P_\infty \sim (p - p_c)^\beta$, consistent with Equation (2). Here, the value of the critical exponent, β, was based on the value of, α as follows:

$$\beta = \begin{cases} 1/(3-\alpha) & for\, 2 < \alpha < 3, \\ 1/(\alpha-3) & for\, 3 < \alpha < 4, \\ 1 & for\, \alpha > 4. \end{cases} \qquad (5)$$

This result confirmed the previous finding with regard to the behavior of the threshold, p_c, in a scale-free network. However, the value of β and the order of the transition and was found to differ depending on the value of α.[9] Equivalent results to Equation (5) were also derived independently by researchers studying percolation in epidemiologic models of disease spread (Pastor-Satorras and Vespignani, 2001a), as discussed below.

The values for the critical exponents in Equation (3) and Equation (4) were determined for scale-free networks and were also found to behave as power laws and exhibit scale-invariant properties. For Equation (3), the number of clusters (i.e., subgraphs) of size s, or n_s, was found to be given by $n_s \sim s^\tau e^{-s/s*}$ as p approaches p_c (Cohen, ben-Avraham, and Havlin, 2002; Newman, Strogatz, Watts, 2001). As in the case of Equation (2), the value of the critical exponent, τ, was also dependent on α. For example, for $2 < \alpha < 3$, the range of α believed to apply to the Internet, it was calculated that $\tau = (2\alpha - 3) / (\alpha - 2)$. At the percolation threshold, the value of $s*$ was approximated by the expression $s* \sim |p - p_c|^\delta$, which was found to diverge at the threshold. Again, for $2 < \alpha < 3$, δ was given by $\delta = (3 - \alpha) / (\alpha - 2)$. For a complete listing of values of δ and τ at additional ranges of α, see (Cohen, ben-Avraham, and Havlin, 2002).

[9] The parameter exponent β attained its usual value (a second-order transition occured) only for $\alpha > 4$. The authors argued for $\alpha < 4$, "the percolation transition is higher than 2nd-order: in particular, for $3 + 1 / n - 1 < \alpha < 3 + 1 / n - 2$, the transition is of the *nth* order" (Cohen, ben-Avraham, and Havlin, 2002), and that "an *infinite-order* phase transition exists at $\alpha = 3$ for growing networks of the Albert-Barabási model". (An infinite-order phase transition is implied when all derivatives go to 0 at the critical point (Dorogovtsev, Goltsev, and Mendes, 2008).) For $\alpha < 3$, the transition was found to occur at a *vanishing threshold*, $p_c = 0$ (Cohen, ben-Avraham, and Havlin., 2002).

In (Cohen et al., 2001), the average distance between two sites in the giant connected component was also estimated, given the correlation length, $\xi \sim |p - pc|-\gamma$, as in Equation (4), where $\gamma = 1$. The average distance, d, was found to differ as follows for: (1) diluted graphs at criticality; and (2) highly-connected graphs above criticality, which include scale-free graphs. In the first case (1), the relationship between d and the giant connected component's size, M, (the number of sites contained within the giant connected component) was given by d ~ M1/2 (1). In case (2), d was found to be proportional to log(kavvg)N, where (kavg) was the average connectivity, and N is the number of sites. The result was interpreted to mean that distances between sites grew dramatically in a disintegrating giant connected component as scale-free networks came under attack. In this event as pc was approached, communications would be impeded even before the giant connected component disintegrated. The interested reader can see (Cohen, ben-Avraham, and Havlin, 2002) for additional derivations regarding critical exponents.

3.3 Summary

The work on percolation theory provided one possible theoretical foundation for study of catastrophic events as phase transitions in distributed communication networks. Percolation theory concepts provided an explanation for how phase transitions occur in communication networks and how a network might transition into a failed state. Further, the giant connected component provided a basis for measuring the magnitude of the effects of the change. Using this basis, researchers developed expressions for determining the value of important quantities for predicting the percolation threshold and the growth of the giant component. Their results generally supported the idea that percolation transitions in random graph models of networks were continuous, though deviations from this conclusion were possible, as noted above. In addition, the work of the percolation theory researchers provided a basis for investigating phase transitions in other graph structures that were relevant to communication networks, most notably subgraphs (Corominas-Murtra, 2010). Recently in (Buldyrev et al., 2010; Li et al., 2012; Zhao, Zhou, and Liu, 2013; Hu et al., 2011; Hu et al., 2013), percolation transitions triggered by attacks on multiple interconnected networks (e.g., interconnected Internet and power grids) were studied, and this is an area of growing interest.

However, there were shortcomings. Perhaps of greatest concern was that the random graph models studied by researchers in this group were highly abstract and did not reflect the topology of the Internet. This was the case even though simulation models used were quite large. The models studied by these researchers were without exception uniformly of one topology; most often: that of a scale-free network. In contrast, actual Internet topology is believed to be far more heterogeneous. In particular, it has been shown that the Internet consists of service provider-based ASs that are arrayed in interconnected hierarchies, where each AS hierarchy is composed of a distinguishable subnetwork (Willinger et al., 2002; Zhou and Mondragón, 2004a; Alderson and Willinger, 2005; Wu et al., 2007; Oliveira et al., 2008) (see Section 7). The differences between the scale-free network topologies used in research models and the topology of the Internet were pointed out in (Willinger et al., 2002), who argued that the Internet, while having scale-free connectivity overall, was engineered so that high-degree hub sites existed primarily on the periphery. In addition, also as pointed out by (Willinger et al., 2002), computer model studies based on random graph models generally did not include widely-used Internet routing and congestion control protocols, which would likely impact the spreading processes. For these reasons, the attempts of researchers who studied percolation of random graphs to directly use their results to evaluate Internet reliability may have proved to be of limited value in the long run.

4. Transitions to Catastrophic States from the Epidemiologic Point of View

A second approach, also based in part on percolation theory, focuses on the spread of disease in a network. Originating in the epidemiologic research community with the spread of disease in a human population (Kermack and McKendrick, 1927), the approach influenced epidemiologists, biologists, mathematicians, and computer scientists. In this section, the epidemiological approach combines models of disease spread with percolation theory to analyze the spread of undesirable agents in communication networks (i.e., the Internet and WWW), most notably, the spread of computer viruses (Pastor-Satorras and Vespignani, 2001a; Moreno, Pastor-Satorras and Vespignani, 2002; Ganesh, Massoulié, and Towsley, 2005; Zou, Towsley, and Gong, 2007; Borgs et al., 2010; and others).

As in the percolation theory approach, communication networks in this section are modeled using random graphs. Both finite and infinite graphs are again considered. In the epidemiologic approach, a disease agent spreads through the network, occupying increasing numbers of sites (in site percolation), until a critical threshold is reached and a percolation transition takes place. The result is a catastrophic event, in which the giant connected component represents the proportion of infected sites. Among the different epidemiologic models of disease spread (Dorogovtsev, Goltsev, and Mendes, 2008), the Susceptible-Infected-Susceptible (SIS) model was most often used for communication networks. The SIS model is widely-used in the epidemiologic literature and is related to the well-known contact model of disease spread, which originated in (Harris, 1974) and was extended by (Bramson, Durrett, and Schonmann, 1991; Liggett 1992; Pemantle, 1992; and others. For surveys of the contact process, see Liggett 1999; Easley and Kleinberg, 2010).

In the SIS model of (Pastor-Satorras and Vespignani, 2001a), the rate spreading was given by $\lambda = v / \sigma$, where v was the susceptibility of a site to infection, if linked to another infected site, and σ was the rate at which sites were "cured". Once cured, a site could become susceptible again. The variable, λ, was the control parameter. If λ exceeded a critical value, λ_c, the proportion of infected sites, ρ, underwent a phase transition to a *stationary endemic* level, or *stationary epidemic* state, after which it remained constant over time (unless counteracted). In (Pastor-Satorras and Vespignani, 2001a), ρ was used as the order parameter, analogous to P_∞ in the percolation studies of the previous section. Figure 4 illustrates the concept.

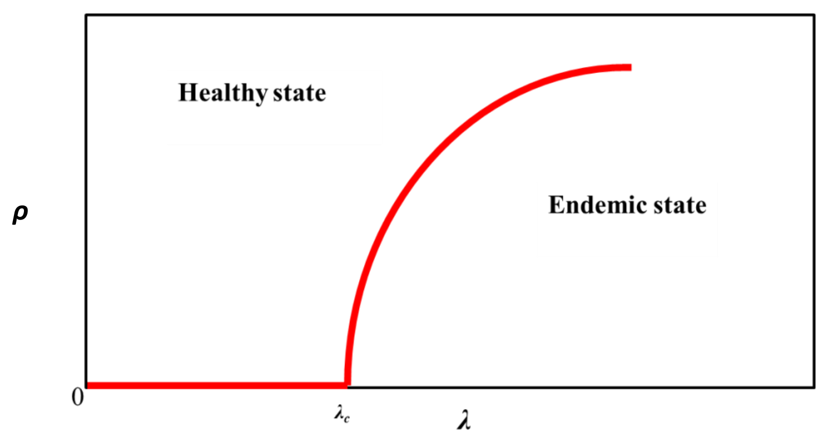


Figure 4. A conceptual representation of the phase transition to the endemic state in an SIS model, or stationary epidemic state, in which the proportion of infected sites, p, is the order parameter. The value of p undergoes a phase transition to the stationary epidemic level at the epidemic threshold, λ_c, while below λ_c, the model is shown to be in a healthy state. Here, the network is assumed to have finite dimensions, as would be the case in a real-world system. The figure is adapted from (Pastor-Satorras and Vespignani, 2003).

Note that in the epidemiologic work, the observed percolation transition did not signify the appearance, or dissolution, of an operational network, but instead, the appearance of a stationary epidemic state caused by the effects of a computer virus. While the SIS model is most frequently used in communication networks and is the focus of this section, other models of disease spread, most notably the Susceptible-Infected-Removed (SIR) model, were also studied and will be touched upon.

Because of the very large amount of literature on epidemic spread, this survey is limited to those works that directly concern distributed communication systems, i.e., the Internet and the World Wide Web. For the context, and an overview, of epidemic spread, see Durrett 2010. In many of these works, the phase transition to the stationary epidemic state was held to be a continuous phase transition (Pastor-Satorras and Vespignani, 2001a; Pastor-Satorras and Vespignani, 2003), as is the case in the more general literature on disease spread (Grassberger, 1983). The phase transition to the stationary epidemic level was also considered a percolation transition by many epidemiologic researchers, including (Grassberger, 1983; Moore and Newman, 2000; Pastor-Satorras and Vespignani, 2001a; Vazquez and Moreno, 2003; Zou, Towsley, and Gong, 2007). As in the percolation theory approach, researchers using the epidemiologic approach analytically sought to derive estimates of important quantities, focusing on λ_c and the dynamic behavior of the order parameter, ρ (recall Table 2.1). Work by (Dorogovtsev, Goltsev, and Mendes, 2008), showed that expressions for estimating some important quantities relating to the percolation transition also applied to the transition to the stationary epidemic state, and these parallel findings are discussed below. In addition, like the percolation theory researchers, researchers using the epidemiologic approach also sought to demonstrate the existence of scale-invariant behaviors in network models based of random graphs (recall Table 2.2).

4.1 Key Papers Using Percolation Theory and the SIS Model

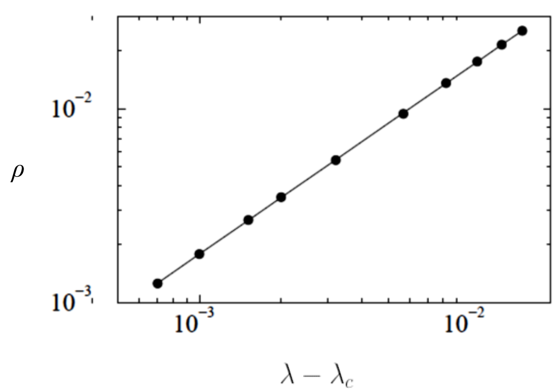
The work of these researchers is probably the most heavily cited among those using the epidemiological approach. A key aspect of this work is the establishment of bounds for key quantities in λ_c and ρ (and the practical absence of such bounds for λ_c in scale-free networks), and congruence with the findings of the percolation theory researchers with respect to power-law relationships (Equations 2–4) and other phenomena discussed in Section 3. In (Pastor-Satorras and Vespignani, 2001a), the SIS model of infection spread was studied in a Watts-Strogatz small-world network[10] using mean-field reaction equations in a stationary system. Assuming a stable average connectivity, k, and an ensemble average $\langle k \rangle$, the epidemic threshold, λ_c, was found to be $\lambda_c = 1 / \langle k \rangle$. For $\lambda < \lambda_c$, the proportion of infected nodes, $\rho(t)$, was shown to decay a well-defined rate. At and near λ_c, the network behaved in agreement with Equation (2), and was found to exhibit scale-invariance, or self-similarity. Here, the relationship between ρ and λ, was given by $\rho \sim |\lambda - \lambda_c|^{\beta}$, where $\beta = 0.97 \pm 0.04$, which corresponded to a slope of +1 on a log-log scale. Similarly, a log–log relationship was found between the proportion of infected sites, ρ, and $\lambda >$

---

[10] A Watts-Strogatz small-world network (Watts and Strogatz, 1998) is a ring topology network consisting of n vertices each interconnected with k neighbors, but in which a subset of vertices, chosen with a probability $0 < p < 1$, are rewired so that they are linked to more distant vertices in the ring. This results in a network that is both highly clustered and has short path lengths. Specifically, starting with a ring of n vertices, each vertex is connected to its k nearest neighbors by undirected edges. A vertex linked to its nearest neighbor is then selected, and the link is rewired to another vertex chosen uniformly at random, with a probability p from the other vertices in the ring. The process is repeated by moving around the ring, considering each vertex in turn, until the entire ring is circled. The circular rewiring algorithm is then repeated for edges that connect vertices to their second-nearest neighbors. As there are $nk / 2$ edges in the entire graph, the circular rewiring algorithm repeats $k / 2$ times. For another, and slightly earlier, version of the small-world model, see (Bollobás and Chung, 1988).

λ_c. In Figure 5a, simulation results are shown which illustrate that phase transitions closely match mean-field predictions.

As with percolation other studies (Erdős and Rényi, 1961; Bollobás, 1984; Cohen et al., 2000; ben-Avraham and Havlin, 2000; Bollobás and Riordan, 2006; and Dorogovtsev, Goltsev, and Mendes, 2008), the application of the SIS spreading process to scale-free networks was of great interest. In (Pastor-Satorras and Vespignani, 2001a), a Barabási-Albert scale-free network was studied in which the connectivity distribution was given by $P(k) \sim k^{-3}$. Using a mean-field theory approach, the proportion of infected sites, ρ, was found to be given by the expression, $\rho = -e^{1/m\lambda}$, where m represented the minimum degree of any site in the network. Using the intuitive argument that increasing the connectivity, k, among sites (and presumably also increasing the value of m) reduced the epidemic threshold, the authors concluded that an epidemic threshold was effectively 0. That is, if the model was theoretically allowed to have infinite connectivity, $k \rightarrow \infty$, then $\lambda_c \rightarrow 0$. In a simulation model of a network with finite dimensions, infections were shown to reach a stationary epidemic level, i.e., ρ would be greater than 0, for very small values of λ (see Figure 5b). These results were extended to a "generalized" scale-free network, in which the connectivity distribution was given by $P(k) = (1 + \alpha)m^{(1+\alpha)}k^{(-2-\alpha)}$, where m is the minimum connectivity for any site. Assuming a network that could theoretically have infinite connectivity, for $\alpha > 1$, the epidemic threshold was given by $\lambda_c = (\alpha - 1) / m\alpha$. However, for $0 < \alpha < 1$, the threshold was found to be absent, as was the case for the epidemic threshold in the Barabási-Albert scale-free network discussed above, and as was largely the case for percolation threshold in scale-free networks described in Section 3 (Cohen et al., 2001). The proportion of infected sites, ρ, was found to be determined by $\rho \sim (\lambda - \lambda_c)^{\beta}$ for $\lambda > \lambda_c$, which again is the power-law relationship of Equation (2). For the case of the "generalized" scale-free network, it was noticed by (Dorogovtsev, Goltsev, and Mendes, 2008) that the epidemic threshold, λ_c, was related to the critical exponent, β, in a manner consistent with Equation (5), which was used by (Cohen et al., 2001) to estimate the proportion of sites in the giant connected component, i.e., the order parameter in an observed percolation transition. Careful examination of the results of (Cohen et al., 2001) and (Pastor-Satorras and Vespignani 2001a) confirms the similarity of the behavior of the giant connected component in scale-free networks for similar values of the exponent β.



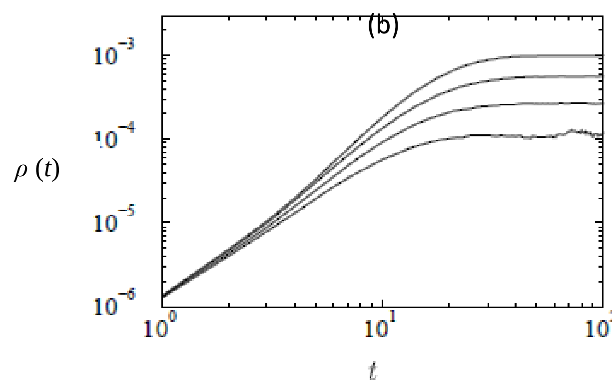


Figure 5 (a,b). (a) Proportion of infected sites ρ as a function of λ in the Watts-Strogatz small-world, where λ_c = 0.1643, ±0.01. Here, the line is a fit to the form $\beta \sim (\lambda - \lambda_c)^{\beta}$, with an exponent β = 0.97 ± 0.04. These simulation results were found to closely match analytical results obtained using mean-field equations. (b) Proportion of infected sites $p\ (t)$ as a function of time in a simulated Barabási-Albert scale-free network with 10^6 sites. The spreading rate, λ, ranges from λ = 0.05 to 0.065 (bottom to top). Figures are from (Pastor-Satorras and Vespignani, 2001a).

These results were extended to the study of the SIS infection spread model in "bounded" scale-free networks having a finite number of sites with maximum connectivity (Pastor-Satorras and Vespignani, 2002a). Here, the epidemic threshold was found to be given by $\lambda_c = \langle k \rangle / \langle k^2 \rangle$, which was generally smaller than the epidemic threshold for the SIS Watts-Strogatz small world network. In (Moreno, Pastor-Satorras and Vespignani, 2002), the spread of infection according to the SIR model in random networks was studied, and here the scale-free model was also confirmed to have a low epidemic threshold, which was equivalent to the epidemic threshold for the SIS model. Analogously to (Cohen et al., 2000; Cohen et al., 2001) who surmised that the Internet was vulnerable to attacks on hubs, the findings of (Pastor-Satorras and Vespignani, 2001a; Pastor-Satorras and Vespignani, 2002a; Moreno, Pastor-Satorras and Vespignani, 2002) were also thought to imply that the Internet was vulnerable to spread of viruses. However, in a follow-up study that included analysis of historical Internet virus data, it was also pointed out that there was an "exponentially small prevalence" of infections reaching a stationary epidemic level "for a wide range of spreading rates" (Pastor-Satorras and Vespignani, 2001b). In (Boguñá, Pastor-Satorras, and Vespignani, 2003), the absence of an epidemic threshold was also confirmed for networks having degree correlations, i.e., networks in which connectivity between sites depends upon the degree of the sites involved, a condition which was thought to be relevant to disease spread. However, the findings relating to Internet vulnerability were, as in the case of the percolation theory researchers, found of limited applicability as knowledge of actual Internet topologies and protocols came to be better understood (Alderson and Willinger, 2005).

4.2 An Approach based on Eigenanalysis of, or Using Spectral Methods on, the Connected Graph

A different method of estimating the epidemic threshold for SIS disease spread in computer networks with directed links was discovered by (Wang et al., 2003), which was based on eigenanalysis of, or use of spectral methods on, a connected graph, $G = (N, L)$, in which N was the number of sites and L was the set of edges between them (as defined in Section 1.1). They proposed a discrete-time model that considered a virus with a birth rate, ω, on an edge to an infected site, which if successfully transmitted would be cured at the infected site at the rate, δ. For **A**, an adjacency matrix of a graph G, having a homogeneous, star[11], infinite power-law, or finite power-law topology, the authors presented a proof that an epidemic threshold existed at T, such that $T = 1 / \lambda_{1,\mathbf{A}}$, where $\lambda_{1,\mathbf{A}}$ was the largest eigenvalue of **A**. If $\omega / \delta < 1 / \lambda_{1,\mathbf{A}}$, the infection decayed and died out over time. An epidemic was prevented if $\delta > \delta_c = \beta \times \lambda_{1,\mathbf{A}}$. For a scale-free network having infinite size, the epidemic threshold was found to be zero, which were essentially in agreement with the conclusions of Pastor-Satorras, Vespignani, and their associates, as well as conclusions of the percolation theory researchers. For a finite power law network, the epidemic threshold was found to exist at $T = 1 / \lambda_{1,\mathbf{A}}$, where $\lambda_{1,\mathbf{A}}$ is the *first* eigenvalue of **A**. The authors executed simulations on the network types they studied (including those of Barabási and Albert and Erdős and Rényi) and concluded that their method was more accurate than that of (Pastor-Satorras and Vespignani, 2001a). They argued that their method was a general analytical model that applied to arbitrary graphs (including Erdős-Rényi random graphs), which considered the impact of topologies but was not limited by it. A subsequent paper by (Van Mieghem, Omic, and Kooij, 2009) extended the work of (Wang et al., 2003) by providing additional proofs confirming the existence of a well-defined threshold at $T = 1 / \lambda_{1,\mathbf{A}}$, but that this result was accurate only when the spreading rate is below the epidemic threshold. The paper by (Van Mieghem, Omic, and Kooij, 2009) discussed and analyzed two Markov models of virus spreading in networks: the exact 2^N–state Markov chain and their newer N-Intertwined Markov model, where for the latter, the largest eigenvalue was also shown to define the

---

[11] The authors also showed, more precisely, that $T = 1/\surd^{1/2}$, where d is the number of satellites in the case of a star network.

epidemic threshold. The accuracy of the two models was studied using simulation of small-sized networks. The *N*-Intertwined model was found to be more accurate below the epidemic threshold, but the models were more equal in accuracy as *N* grew larger. The paper by (Van Mieghem, Omic, and Kooij, 2009) considered the effect of the structure of the network on epidemic spreading as did (Ganesh, Massoulié, and Towsley, 2005; Zou, Towsley, and Gong, 2007), discussed further below.

4.3 Incorporating Realistic Characteristics and Other Epidemiologically-based Approaches

The work of (Zou, Towsley, and Gong, 2007) offered still another perspective on the spread of viruses in communication networks and drove the study of phase transitions into the direction of incorporating real-world characteristics into models. This study focused on email worms, a class of viruses which spread over logical networks defined by email address relationships, which constitute a subset of the overall topology determined by network connectivity. Accordingly, the authors argued that epidemic models based on network topology were inaccurate in studying the propagation of email worms. Further, they argued that topological epidemic models, such as those of Pastor-Satorras and their associates, which were analyzed using “differential equations”, overestimated epidemic spreading. Further, they argued that the work of (Wang et al., 2003) was also limited because their equations only represented the final stationary epidemic state and not the evolution of the infection. Therefore, they developed a SIR-based model, based on email address relationships, which simulated email worm spread. The model accounted for realistic behaviors of email users, including such factors as email checking time and the probability of opening an email attachment. They used the SIR (rather than the SIS model) because they assumed once a virus was found and eliminated, the host was unlikely to be infected again.

Using a simulation model, (Zou, Towsley, and Gong, 2007) studied the spread of email worms in scale-free random graph email address topologies, as well as homogenous random graph and small-world topologies. As in the case of previous works surveyed here, they observed the transition of the email address network to a stationary epidemic state. Their findings indicated that Internet email networks followed a heavy-tailed connectivity degree distribution, suggesting that such networks may have a scale-free topology. To counter email worm infections, they developed a strategy of immunizing selected users, who were known to function as hubs and who sent emails to large numbers of other users. Here, they computed thresholds for the number of high-degree hub users that needed to be immunized to prevent infection and characterized the immunization process as being related to a “percolation problem”. They found that email worms spread more quickly on scale-free networks than on small world topologies or random graphs, but that immunization defense was also more effective. The work of (Zou, Towsley, and Gong, 2007) showed the importance of studying specialized simulation models that are based on real-world characteristics.

Like (Zou, Towsley, and Gong, 2007), a few researchers also studied methods to counteract infection spread. In (Joo and Lebowitz, 2004), the SIS spread process was modified to simulate “saturation” effects and the impact of firewalls on viruses, in which transmission of the infection through links depended on such factors as contact time or bond strength. In (Schwarzkopf, Rakos, and Mukamel, 2010), it was argued that rewiring the network (i.e., changing links between sites) had the effect of raising the threshold for transition to a stationary epidemic state in a scale-free network with $\alpha > 3$ (for scale-free networks in which $2 < \alpha < 3$, recall that it was found there is no effective threshold). For homogenous networks, rewiring had little effect.

In (Borgs et al., 2010), immunization procedures and the distribution of antidote, were studied for various topologies including those having a power-law distribution such as the WWW, using the contact process to model the spread of viruses and worms. Here, the total amount of antidote needed, the distribution of the antidote, and the duration of the epidemic, were stated in theorems on finite, undirected graphs. Immunization procedures to counteract the effects of disease spread were also studied in (Pastor-Satorras and Vespignani, 2002b).

Other papers also considered the effect of network structure of virus spreading, including (Ganesh, Massoulié, and Towsley, 2005; Van Mieghem, Omic, and Kooij, 2009). In particular, (Ganesh, Massoulié, and Towsley, 2005) used the model of (Wang et al., 2003) to show that topology affected spreading in a variety of networks having differing topology types. These topology types included using power-law random graphs to model the AS-structure of the Internet and linking Erdӧs-Rényi graphs to AS graphs. They went on to argue that Border Gateway Protocol (BGP) routers (discussed in Coffman et al., 2002), belonging to the top-level AS systems, formed a completely connected graph (an undirected graph in which every pair of distinct vertices is connected by a unique edge). They discussed thresholds for different types of network topologies and went on to show how topology affected the spread of an epidemic. They presented a preliminary discussion of conditions under which epidemics either died out quickly, more slowly, or persisted in a variety of topologies.

Researchers using the epidemiological approach also contributed to the *bootstrap percolation* model (Scalia-Tomba, 1985; Ball and Britton, 2005; Janson et al., 2012), which although more abstract, is relevant to the spread of viruses through communication networks and is closely related to the SIR model. In the bootstrap model based on a random graph, sites are initially occupied randomly with probability p. At each subsequent time step, sites become infected if they have at least m infected neighbors. Sites stay infected, once they become infected. The process repeats until no new sites become infected. In these studies, the bounds of the bootstrap percolation transition were explored in a variety of finite and infinite structures (including graphs). For instance in (Janson et al., 2012), researchers either varied the number of sites initially occupied, or varied the probability that an edge exists between two nodes, in an Erdös-Rényi random graph. They concluded bounds on the number of sites that were occupied at the end of a process. In particular, they were concerned with how the parameters needed to be varied to infect all, or nearly all, the sites—and the time needed to achieve this result. In (Berger et al., 2005), the spread of epidemics across the Internet was analyzed using the contact process on scale-free graphs in preferential attachment models. Assuming a rate which an infected site would communicate a virus to healthy neighbors, the probabilities that the virus would develop into an epidemic were derived.

4.4 Summary

Like the percolation theory researchers, the work of researchers using the epidemiological approach also contributed to the theoretical foundations for study of observed percolation phase transitions in communication networks, with directed focus on phase transitions caused by spreading of viruses and related agents. Like the previous group, these researchers were able to extend their theoretical knowledge base to include formulae for calculating important quantities related to the percolation transition. An important part of their work was the establishment of bounds for key quantities ρ, and λ_c, and the demonstration of low resistance to disease spread in scale-free networks. Another important aspect of their work was agreement with the findings of the percolation theory researchers with respect to power-law relationships, self-similarity, and other phenomena discussed in Section 3. However, as with percolation theory researchers in the first group, most of the random graph models studied by this group, though substantial in size, were also highly abstract and did not reflect topologies found in the Internet. This appears to be the case, even though some attempts were made to study virus spread using real-world data and under assumed real-world circumstances. As before, the models studied were of one uniform topology (excepting the researchers who chose to study the effect of network structure on virus spreading), and did not reflect the characteristics of the far more heterogeneous Internet, which is composed of sub-networks with different topological types. In addition, they also did not consider factors such as routing policy and congestion control techniques. Hence, attempts to use their findings to evaluate vulnerability of the Internet also did not produce durable results, with respect to the Internet.

5. Studies of Percolation Transitions Caused by Cascades

The focus of this approach is the study of cascades, sometimes also called avalanches, in communication networks. As in the previous two approaches, a random graph of a particular type is generally used to model the network. In the cascading process, a quantitative property initially occupies a small set of sites (possibly as few as one) from which it spreads, or *cascades*, to adjacent sites. Sites that are occupied by this property are assumed to be inoperable and therefore disconnected from the rest of the network. Each site has a quantitatively expressed capability to resist the spread of the cascading property. If a site (or sites) is occupied by the quantitative property whose value exceeds the threshold resistance capability of an adjacent site, then the cascading property spreads to, and occupies, the adjacent site. In this manner, the cascade continues from site to site. If the number of occupied sites becomes sufficiently large, it will result in the breakup of the giant connected component of unoccupied sites, an event which represents the destruction of network connectivity and a phase transition to a global inoperable state.

The cascade studies described in this section utilize percolation concepts to a significant extent, and it is possible to divide these studies into two subcategories in this respect. Researchers in one subcategory, like researchers who used the approaches described in the preceding sections, used percolation models as a basis to study the properties of global phase transitions in communication networks. In the second subcategory, percolation theory was not explicitly used to infer or interpret results. Instead here, researchers used simulation models to conduct empirical studies of cascades, and supported their findings in some cases with analytical studies. Both subcategories of researchers determined critical thresholds for cascades and estimated the change in the size of the largest connected component which was formed as the cascade proceeded (recall Table 2.1), though these quantities were determined analytically by the former group and empirically by the latter. Communication networks are modeled using finite graph topologies, with distance used as a weight factor for links in a few cases.

Also of interest is the property whose spread occupied sites as the cascades proceeded. In the cascade studies grounded more directly in percolation models, the property was allowed to be generic and unspecified. However, in the second category of cascade studies, the property was load. In this respect, a relationship exists between the works in this section, which study load cascades, and the papers in the next section, which discuss phase transitions caused by congestion. The distinction between the two is that in the former, the focus is on the mechanics by which the cascade spreads and causes the breakup of the giant connected component that represents network connectivity (recall Section 3). In the latter (as we shall see), this focus is almost totally absent, and the works in the next section are more concerned with the properties of network congestion than with the manifestation of the phase transition itself. The cascade studies of the spread of overload in networks parallel cascade studies in power grids (Carreras et al., 2002; Newman et al., 2011), unlike the congestion studies described in the next section, which do not. Finally, the reader may discern that there is significant overlap between the cascade concept and epidemiologic spread, particularly with respect to the SIR model of epidemiologic spread. These two processes can perhaps best be distinguished by the definition of the spreading process, the model used to represent the spreading or cascading process, and the type of real-world phenomena that is being modeled. Arguably, these categories could potentially be merged. The exploration of the similarities between these two approaches remains a topic of further work.

5.1 Cascade Studies based on, or Related to, Percolation Theory

In (Watts, 2002), Watts developed a percolation-based model of cascade spread in random graphs with different degree distributions. In this model, a site became occupied by the cascading property if at least a threshold fraction, ϕ, of its k (connected) neighbors were also occupied. A site having sufficiently low ϕ or high k could be deemed vulnerable to occupation. To execute the model, a random value of ϕ was chosen for each site, a small population of occupied sites was introduced, and the cascade was initiated. Using a generating function approach (Newman, Strogatz, and Watts, 2001), Watts computed the distribution of vulnerable sites having a particular degree and estimated the distribution of clusters of vulnerable sites. He then used these results to calculate the critical number of vulnerable sites, z, where $z = \langle k \rangle$, needed for the cascade to result in the formation of a giant connected component composed of occupied sites. What Watts observed suggested that there were two types of phase transitions at z, depending on connectivity of a site: a more frequent, continuous, second-order phase transition and a rare, discontinuous, first-order phase transition. When connectivity was low, the distribution of cascades followed a power law distribution and the phase transition was continuous, second order. When connectivity was high, most clusters of occupied sites were small, and the cascade often died out without undergoing a global phase transition. However, under conditions of high connectivity and sufficient vulnerability, a much larger cascade could occur that resulted in a discontinuous, first-order transition to a very large giant cluster. In (Watts, 2002), Watts also studied the complicating effect of system heterogeneity, in the form of site degree distribution, and observed mixed results. He found that increases and decreases in heterogeneity both increased and decreased the likelihood of global cascades. In (Gleeson and Cahalane, 2007), the work of Watts was extended to show additional dependencies between the size of global cascades and the size of the initial seed. This follow-on work discovered conditions under which continuous and discontinuous phase transitions may occur depending on connectivity and the value of z. Additional studies of cascades, also based on the percolation model, were conducted by (Gleeson, 2008; Hackett, Melnik, and Gleeson, 2011). In (Gleeson, 2008), the study of cascades was extended to networks having degree correlations. Also the effect of topological structure continued to be studied in (Gleeson, 2008), including k-core sizes, where a k core of a network was considered "the largest subgraph whose nodes have degree at least k".

The work of (Samuelsson and Socolar, 2006) also used percolation theory to study cascades in communication networks, where the network was modeled as a random Boolean network. As in other random networks, a random Boolean network was formed by randomly connecting pairs of sites. Each site had a binary state, 0 or 1. The value 0 signified the site was undamaged, while 1 indicated it was damaged. The cascade propagated to a site (i.e., damaged a site) on the basis of a Boolean rule which examined if (i.e., used as inputs) the site's neighboring sites were also damaged, i.e., had a value of 1. This work was distinguished from other cascade studies in that it focused on the concept of *exhaustive percolation*, in which cascades affected all sites in an observed network. Initially, a small fraction of sites was selected to be damaged and begin the cascade process. The authors derived analytical expressions to determine the probability that the percolation event was exhaustive and, in the event that it was not, the probability distribution of the number of unoccupied sites. Key to their formalism was the concept of the *Unordered Binary Avalanche*. This concept combined the concepts of Random Boolean Networks with those of the probability of nodes being damaged. Their exhaustive percolation formalism stated in (Samuelsson and Socolar, 2006) was extended to a broader class of networks (including scale-free networks). In this way, they hoped to explain the spread of viruses and other related phenomena.

5.2 Empirical Studies of Cascades

A number of researchers produced empirical studies of cascading processes in communication networks, which when sufficiently large, caused the break-up of network connectivity. They observed the breakup of connected structures, where the structures resembled a giant connected component. In contrast to (Watts, 2002), generally there was no attempt to characterize the phase transition order or to identify self-similar patterns. Instead, the focus seemed to be on estimating the critical value, defined in terms of resistance to attack, needed by each node to thwart a cascade—in other words to prevent a network from falling apart in response to a cascade.

In work by (Motter and Lai, 2002), cascades were initiated by attacks on individual sites in a simulation model of a scale-free network, where $P_k \sim k^{-\alpha}$. Here, an attack on a site resulted in the shifting of its load to neighboring sites, which can become congested and fail. The failure of a neighboring site could cause load to be diverted to other sites, which also became overloaded and failed, and so forth. Individual site capacity was defined by $C_i = (1 + \psi_i)L_i$, where, $\psi_i \geq 0$, was a *tolerance* parameter, and L_i was the initial load on the site. The success of the attack was measured by $G = N' / N$, where N is the number of sites in the largest network subcomponent before the cascade, and N' was the largest connected network subcomponent afterwards. Like some Percolation theory researchers, the results of the study indicated that targeted attacks on sites with larger loads (including hubs with many links) led to significant cascades[12] in finite networks. However, whether the method of load diversion to neighboring sites is accurately reflective of real-world networks is a matter of debate. It appears that the model of (Motter and Lai, 2002) was intended to be generalized for electrical grids as well (Carreras et al., 2002).

The vulnerability of highly connected sites (hubs) and high-traffic sites in models of scale-free graphs was confirmed in a follow-up study (Lai, Motter, and Nishikawa, 2004). Using the model of (Motter and Lai, 2002) as a basis, (Zhao, Park, and Lai, 2004) executed a study on cascading failures in a Barabási - Albert scale-free (SF) network and observed evidence of a phase transition in a finite network. Here, an analytic solution was provided to estimate the critical value of a tolerance parameter, ψ_c, needed to thwart an attack, where values above ψ_c enabled the attack. Using the same measure of attack success, G, given in (Motter and Lai, 2002), the study showed both analytically and through simulation that for $\psi < \psi_c$, G was positive. However, while the network was said to fall apart and become non-functional when $G \rightarrow 0$, as in (Motter and Lai, 2002), no explicit relationship was drawn between the breakup of the network and the disintegration of the giant connected component as defined in percolation theory. Similarly in (Lai, Motter, and Nishikawa, 2004), as in the studies by (Motter and Lai, 2002; Zhao, Park, and Lai, 2004), the largest connected subcomponent, N, was not explicitly related to a giant connected component as defined in Section 3.1. To predict impending network breakdowns, the authors recommended monitoring of ψ to detect its proximity to ψ_c. The network models studied in (Motter and Lai, 2002; Zhao, Park, and Lai, 2004; Lai, Motter, and Nishikawa, 2004) did not differentiate between hosts and routers and did not consider routing or Internet protocols.

Still using the basic model of (Motter and Lai, 2002), (Lee et al., 2005) studied cascade dynamics in a scale-free random graph model, with the scale-free degree distribution, $P_k \sim k^{-\alpha}$. Here, cascades were triggered by failing selected sites, which again caused the load that passed through the failed sites to be diverted, or detoured, to neighboring sites, whose load was correspondingly increased. If the increased load to a diverted site exceeded its capacity, it too failed, thus propagating the cascade. In this study, cascade dynamics were governed by the *sand pile model* (Bak, Tang, and Wiesenfeld, 1988), which

[12] Note that this study did not draw a strict correlation between load and the degree of the node.

provided a known cascade size distribution based on the size of ψ. Consistent with the sand pile model, cascade distribution in a scale-free network followed a power law, $p_a(s) \sim s^{-\tau}$, where s was the avalanche size and $\tau = \alpha / (\alpha - 1)$ for $2 < \alpha < 3$). The paper by (Lee et al., 2005) made the following findings in a finite network, which were then held to be relevant to potential Internet vulnerability: (1) a critical value $\psi_c = 0.15$ was discovered at which the sand pile power-law model held for the range, $2 < \alpha < 3$; (2) the value of, τ, was found to be relatively stable at $\tau \approx 2.1$ for $2 < \alpha < 3$; (3) for small value of ψ, cascading failure spread over the entire system, but for a large ψ, cascading failure is confined in a small region; (4) at ψ_c, avalanche size distribution follows a power law with an exponent of τ; and (5) vertices having moderate (or *intermediate values*) degree, k, were found to be more vulnerable to failure by cascades. The authors stated: "Accordingly, when a vertex on a branch of tree structure is removed, the giant cluster is divided into two or more components, and the giant cluster size shrinks apparently". Like (Motter and Lai, 2002; Lai, Motter, and Nishikawa, 2004; Zhao, Park, and Lai, 2004), the simulation model used in (Lee et al., 2005) did not differentiate between hosts and routers, did not consider routing or Internet protocols, and its view of load redistribution was in question.

In (Moreno, Gomez, and Pacheco, 2002), a study was conducted "within the framework of percolation theory", which reported the occurrence of cascades in a scale-free network through propagation of overloads. In this study, each site was able to support a finite load, σ, together with a "security threshold", σ_{ith}. To assign the threshold values to each site, a Weibull distribution $P(\sigma_{i\mathrm{th}}) = 1 - e^{-(\sigma_{i\mathrm{th}})^{\rho}}$ was used, where ρ is the Weibull index, which was used to control the degree of threshold disorder in the system (the bigger the value of ρ, the narrower the range of threshold values). If the load applied to a site exceeded the "security threshold", the site failed. A failed site's load then transferred to neighboring operational sites (i.e., sites with links to the non-failed sites), and the process was repeated. The results showed that when a critical load level was reached, a cascade would occur in which a functional network abruptly transitioned to a disintegrated, fragmented network. See Figure 6. Here, a functional network was measured by the size of its "giant component", which was described as the largest connected subcomponent in the network. The results show that as ρ increased, the critical point shifted rightward (also increased). The authors observed that the critical load level for a given value of ρ was independent of network size. They also estimated the distribution of avalanche sizes. In contrast to (Motter and Lai, 2002; Zhao, Park, and Lai, 2004), in which cascades in scale-free networks originated at hub sites, this study found that cascades could begin at random points within the network. Yet even though hubs were not targeted, the probability of hub failure was found to be proportional to connectivity, and this probability increased with the increase in load. The model of (Moreno, Gomez, and Pacheco, 2002) did not differentiate between hosts and routers, did not consider routing or Internet protocols, and was consistent with other studies with respect to load redistribution, but consisted of as many as 10 000 sites.

In (Moreno et al., 2003), Moreno extended his earlier work on cascades caused by load redistribution in a Barabási-Albert scale-free network. The cascade again depended on increasing the average load across the network and overloading selected links. When the load on a link exceeded its capacity, the link was considered congested, and the load was diverted to links departing from the end site of the congested link, thus initiating the cascade. Here, the authors appear to focus more on overloading links than the departing nodes on links. In one of the few studies to address the issue of redistribution of load for overloaded nodes, different formula for initial distribution and alterative rules for re-distribution (upon failure) were tested, and simulations were run for them. The authors studied the transition from a network-wide free phase to a congested phase and the disintegration of a giant component, which was the largest (connected) component of the network. The authors found that above a critical value of the average traffic load, a single failure had a finite probability of triggering a congestion avalanche which destroyed the giant component and thus network connectivity as a whole. This event was distinguished from a percolation transition in that above this critical value, the probability of having a giant component steadily decreases to 0 as load is progressively increased. Finally, at some ceiling load value, this probability reached 1. The authors found that the probability decayed to zero and that there was a wide region of (values) where the initial instability could trigger congestion with a probability (greater than 0).

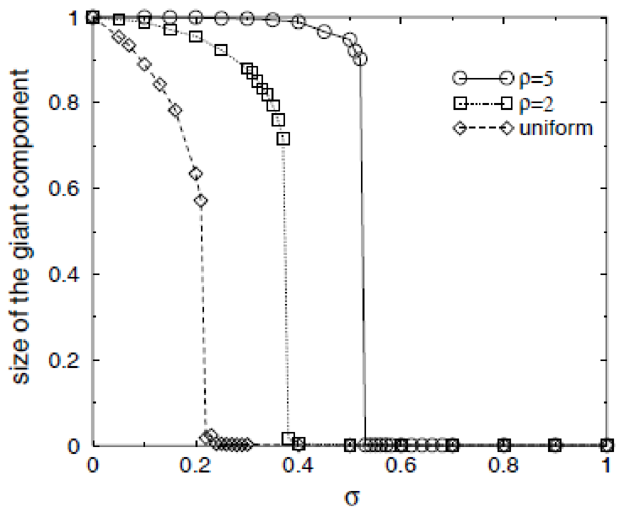


Figure 6. The figure shows the change in the proportion of sites in the giant component for a Barabási-Albert network consisting of 10^5 nodes, in which σ represents the load on the system (in dimensionless units). The values of ρ = 2 and ρ = 5 correspond to different values for the Weibull index parameter, which controls the degree of threshold disorder for two different levels of heterogeneity for the "security threshold" of the sites. As the threshold distribution becomes more homogeneous and the range of threshold values narrows, the critical point shifts rightward; however, the precursory activity in this case is less intensive and the collapse of the network is more catastrophic. The figure is from (Moreno, Gomez, and Pacheco, 2002).

5.3 Other Cascade Studies

The paper of (Coffman et al., 2002) described models of behavior for interacting BGP routers that were interconnected in a "router clique", in which cascading failures lead to phase transitions into states in which all routers in the clique are failed. The definition of *phase transitions* used here was attributed to Erdős and Rényi: "an abrupt change in a global system property", which in this paper is interpreted as a curve with a pronounced vertical jump with no characterization of its order. The event which triggered the phase transition might be an attack, virus, or related occurrence that knocked a certain number of routers out. If the number of routers exceeded a threshold, it triggered a series (wave) of BGP router withdrawal messages, causing the cascade and the resulting phase transition to a state in which all routers are down. In contrast to all other work on phase transitions in communication networks, the cascade process in this study was modeled using both fluid model and a birth-death model. The authors found that "the propensity for phase transitions increases as clique size increases, and additionally also increases as the processing capacity of the routers decreases". They also found that clique size would provide a threshold for the phase transition. Later, using an epidemic model of spreading, (Ganesh, Massoulié, and Towsley, 2005) went on to argue that BGP routers belonging to a top-level AS system formed a completely connected graph. For such a system, they proposed a threshold for cascades, defined essentially as 1 / n, where n is the number of sites in the graph. The authors also proposed other thresholds for systems with different topologies.

In (Holme and Kim, 2002), a Barabási-Albert scale-free network was initially grown through the preferential attachment algorithm of (Barabási and Albert, 1999) until a giant connected component formed. Each site was assigned a random capacity value, and its load was calculated using the *betweenness centrality* metric, which measures the number of shortest paths that run through a site. In this process, the hub sites usually have the highest loads. In the resulting simulations, it was observed that hub sites would become overloaded, triggering cascades that fragmented the network into many small, isolated, chain-like clusters. The authors discussed the implications of these results for Internet reliability. They concluded that even if the process resulted in a slowing down of servers rather than a complete breakdown, a problem could occur in the Internet "if the exponential increase in computer performance stalls, but not the growth of the number of Internet sites". To remedy the problem, (Holme and Kim, 2002) supposed that "overflow control", such as provided by BGP, could be optimized and centers of potential overload could be removed by a careful study of network topology. These conclusions, like those of the percolation and epidemiologic researchers, were somewhat compromised by their reliance on inaccurate models of Internet topologies.

5.4 Summary

As in the previous approaches which were based on percolation theory and models of epidemiologic disease spread, cascade processes were studied in network models that were highly abstract and did not reflect the Internet topology. As before, the models were of one uniform topology, in contrast to the topologically heterogeneous Internet (see Section 7). With few exceptions (e.g., Coffman et al., 2002; Ganesh, Massoulié, and Towsley, 2005), researchers who studied cascading processes also did not consider factors such as routing policy and congestion control. Also except for (Coffman et al., 2002), no model reflected the operation of the transmission control Protocol / Internet protocol (TCP/IP), which could affect the progress of cascades by dampening their flow, rather than redirecting it (see Section 7). Hence, the relevance of their conclusions for the prospects of phase transitions in real-world communication networks was not clear. However, these researchers introduced the study of cascade mechanisms as agents that caused phase transitions in computer network environments. Cascade studies that were based on percolation models provided a consistent view of an underlying process by which cascades caused global phase transitions in networks. Using percolation theory as a basis, in (Watts, 2002; Gleeson and Cahalane, 2007) both continuous and discontinuous phase transitions were observed. However, in the case of (Moreno et al., 2003), we see that a distinction was drawn between percolation transitions due to load and phase transitions induced by cascades. Empirical studies used simulation to provide examples of the breakup of largest network component, which appears to be related to disintegration of giant connected component and the percolation process.

6. Transitions to Catastrophic States Resulting from Increased Load and Congestion

This approach focuses on the study of the effects of congestion in communication networks–in particular, how increasing load caused a phase transition from a network-wide free state to a catastrophic jammed state. Researchers who used this approach developed simulated models of finite communication systems, in which they studied the effects of increasing load and the resulting behavior of the systems as it approached a critical point (Solé and Valverde, 2001; Woolf et al., 2002; Arrowsmith et al., 2004; Echenique, Gomez-Gardenes and Moreno, 2005; Tadic, Rodgers and Thurner, 2007; Lawniczak et al., 2007; Wu, Wang and Yeung, 2008; de Martino et al., 2009; Wang et al., 2009a; Wang et al., 2009b, among others). The approach in most of these studies was primarily empirical, relying strongly on observation of simulations. However, in a number of studies, analysis using mean-field theory equations also played a prominent role. The models used in these studies assumed both two-dimensional lattice and random graph topologies. Here, load, λ, served as the control parameter, defined as the frequency at which packets were injected into their models at selected sites. Selected routing procedures were used to forward packets across sites in a simulated network to their individual destination sites. At higher load levels, packet forwarding was inhibited because queues formed at sites due to local congestion, leading to observable network-wide congestion. At some critical λ_c, increased levels of congestion caused a phase transition from a free state in which packets regularly arrived at their destinations in a timely manner to a congested, or jammed, state, after which network throughput fell dramatically. To identify the phase transition, researchers also observed sharp increases in values of a variety of system variables which reflected congestion, such as mean packet lifetime, average queue size, and number of packets in the network.

Characteristically, percolation theory was not used as a model for the transition to a congested state by most of these researchers. Unlike the previous approaches, the phase transition was not measured by an increase in the occupation of sites by a spreading agent or in the emergence of the giant connected component. Nor was load modeled as a cascading mechanism, as pointed out above; and as seen in other surveys (Boccalettia et al., 2006; Dorogovtsev, Goltsev, and Mendes, 2008), where network congestion was classified separately from cascades. Rather, in this approach, researchers were more focused on measurement of the effects of increasing load and the appearance of resulting phase transitions, together with related phenomena, such as self-similarity and long-range dependence (for which testing methods are discussed in an Appendix). However, the possible relationship between percolation theory, cascades, and the study of congestion is an important topic that will be discussed later. As will be argued, the quantities used by researchers discussed in this section to signify a phase transition to a congested state (e.g., decrease in throughput, increase in packet lifetime, etc.) are possibly outward manifestations of deeper processes that have yet to be identified. This section overviews the work of this class of researchers and discusses their major findings, and points out where the results are inconclusive and more work is needed.

6.1 Simulation Models

The models created by researchers using this approach generally consisted of hundreds, and in some cases thousands, of simulated sites that were configured in a variety of topologies. Generally, individual sites were modeled either as hosts, which originated (and received packets), routers which forwarded them to destinations, or both. For example, a square lattice topology in which sites acted as both hosts and routers was used by (Solé and Valverde, 2001; Woolf et al., 2002; Sarkar et al., 2012), while (Ohira and Sawatari, 1998; Arrowsmith et al., 2004; Mukherjee and Manna, 2005; Lawniczak et al., 2007) modeled a two-dimensional lattice, but differentiated host and router sites. Other studies focused on scale-free networks, in which host and routers were differentiated (Echenique, Gomez-Gardenes and Moreno, 2005; Sarkar et al., 2012), and in which hosts and routers were undifferentiated (Tadic, Rodgers, and Thurner, 2007; Wu, Wang and Yeung, 2008; Wang et al., 2009a; de Martino et al., 2009). Other topologies were also studied, such as randomly connected networks with homogenous degree distributions (de Martino et al., 2009), Erdȍs-Rényi random networks (Wang et al., 2009a), Watts-Strogratz small-world models (Watts and Strogatz, 1998), random boolean networks (Gershenson, 2003), and Barabási-Albert scale-free networks (Wu, Wang and Yeung, 2008). Topologies were primarily randomly generated, except for instance (Echenique, Gomez-Gardenes and Moreno, 2005) where real-world networks were used. Weights on links and sites were not used unless noted.

The routing algorithms used in these efforts appear to be motivated by knowledge of Internet processes, particularly with respect to congestion awareness. For example, (Solé and Valverde, 2001; Woolf et al., 2002; Arrowsmith et al., 2004) used shortest path routing as a primary criteria for determining which site to forward a packet to, with either least congested site as the secondary criteria (Solé and Valverde, 2001), or least used path as the secondary criteria (Woolf et al., 2002; Arrowsmith et al., 2004). A combination of criteria (depending on the researchers) were also used to determine where to route packets in these models, such as shortest path, availability of buffer space at the destination site, and congestion awareness (Echenique, Gomez-Gardenes and Moreno, 2005; Wang et al., 2009a). Congestion aware routing (and modulation) procedures were also modeled in (Lawniczak et al., 2007), and in (de Martino et al., 2009), a feedback mechanism was used to provide congestion information to packet senders. Some researchers used routing procedures based on random choice (Mukherjee and Manna, 2005; Tadic, Rodgers and Thurner, 2007; Wu, Wang and Yeung, 2008; de Martino et al., 2009; Sarkar et al., 2012), while others (Ohira and Sawatari, 1998) compared a deterministic routing algorithm with a probabilistic one. Infinite-sized buffers were assumed, except for (Tadic, Rodgers, and Thurner, 2007; Wu, Wang, and Yeung, 2008; Wang et al., 2009b; Sarkar et al., 2012), where finite buffers sizes with packet dropping procedures for overflow were assumed. In (Wu, Wang, and Yeung, 2008) models were varied to use both infinite and finite buffers. Generally, packets were sent from randomly selected sources to randomly selected receivers.

At the critical load level, λ_c, evidence for a phase transition was observed in the form of significant increases in quantities such as packet lifetime or packet latency, (Solé and Valverde, 2001; Woolf et al., 2002; Arrowsmith et al., 2004), increased number of packets in the system (Solé and Valverde, 2001; Echenique, Gomez-Gardenes and Moreno, 2005; Lawniczak et al., 2007; de Martino et al., 2009; Wang et al., 2009a; Sarkar et al., 2012), increased queue lengths at routers (Solé and Valverde, 2001; Woolf et al., 2002), and increased number of sites with full buffers (Wu, Wang, and Yeung, 2008). Above λ_c, these quantities continued to rise. Steady increases in system throughput were also observed as to $\lambda \rightarrow \lambda_c$, followed by a drop off for $\lambda > \lambda_c$ as the system entered the congested or jammed phase. See Figure 7.

6.2 Characterization of Phase Transitions

A number of researchers provided characterizations indicating whether the phase transitions they observed were discontinuous, first or continuous, second order (recall the discussion of the concept of phase transition order in Section 1.3). In most cases, this characterization was based on observations of the rate of change in measured system quantities that evidenced the phase transitions, rather than theoretical analysis. Most researchers discussed in Section 6.1 appear to have observed second-order phase transitions. As such, as the reader shall see, their results have proven somewhat inconclusive when taken together.

In (Solé and Valverde, 2001), when packet transmission was modeled as an information transfer process, a second-order phase transition was believed to occur at the maximum point of information transfer corresponding to λ_c. Here and previously (Solé et al., 1996), the authors use theoretical phase transitions to determine phase transition order. Second-order phase transitions were described in abstracted systems in (Sarkar et al., 2009; Rykalova, Levitan, and Brower, 2010; Sarkar et al., 2012). In (Echenique, Gomez-Gardenes, and Moreno, 2005), a scale-free network model of 11 174 sites was generated from an Autonomous System Map. Here when a traffic awareness algorithm was used for network routing, a discontinuous, first-order phase transition to a congested state was observed at λ_c. When traffic awareness was absent, phase transitions appeared to be second order. In (Echenique, Gomez-Gardenes, and Moreno, 2005), the determination of the phase transition order was influenced by the steepness of the change in the slope for the order parameter at criticality. Results were also obtained by (de Martino et al., 2009). In this study, a discontinuous transition was found to occur at λ_c in the change in the number of packets over time, in a scale-free network with a heterogeneous degree distribution, in which the routing algorithm used traffic awareness. However, for a random network with a homogenous degree distribution, a continuous phase transition was found to occur. In addition, in a homogenous network, traffic awareness, it seems, did not impact the phase transition. In (de Martino et al., 2009), both a mean-field theory approximation and simulation were used. Here again, the steepness of the change in the order parameter was taken into account in determining whether the transition was continuous or discontinuous. Further, the use of traffic awareness in routing was found to cause λ_c to be higher in both (de Martino et al., 2009) and (Echenique, Gomez-Gardenes, and Moreno, 2005). In (Wu, Wang and Yeung, 2008), the phase transition order in a Barabási-Albert scale-free network was influenced by buffer sizes being finite or infinite. Both first- and second-order phase transitions were observed. Here, the rapidity with which congestion in arose as $\lambda \rightarrow \lambda_c$ was considered when determining phase transition order. No study, it seems, considered different packet sending rates for different types of nodes.

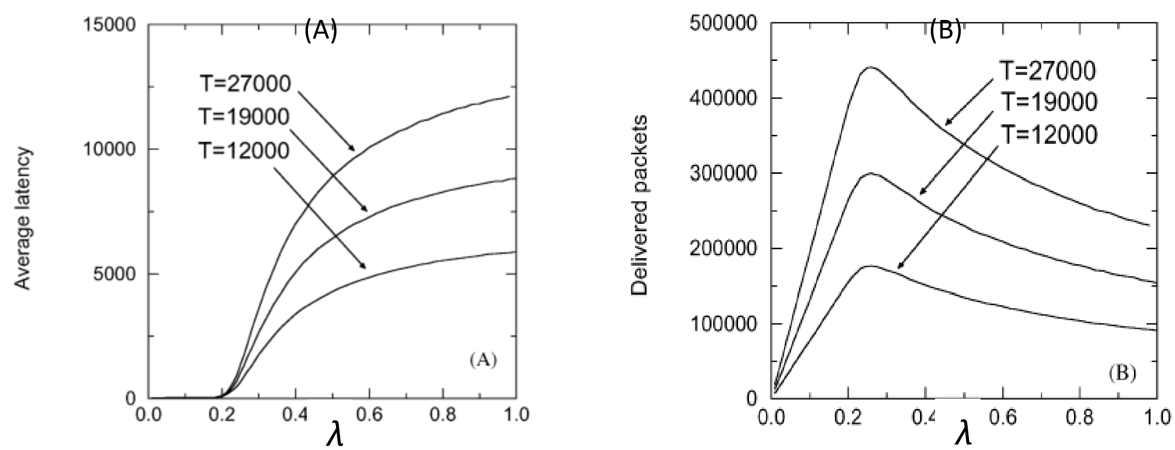


Figure 7. (A) The figure shows a phase transition in network traffic in a 32 x 32 lattice. The average packet latency has been computed over different, increasing intervals of time steps, T, as indicated. (B) As a measure of efficiency, the number of delivered packets has been measured under the same conditions. We can see the optimum at the critical point $\lambda_c \approx 0.2$. Similar results were consistently generated for a number of other papers surveyed in this section. These figures are from (Solé and Valverde, 2001)

While researchers studying congestion provided a more detailed description of system behavior at the critical point than did researchers using the other approaches, clearly more work needs to be done, because the full range of circumstances in which phase transition phenomena may occur is not yet well-understood. This is also the case with related phenomena, such as self-similarity and long-range dependence, which were also examined by researchers who studied phase transitions.

6.3 Findings Relating to Self-Similarity and Long-Range Dependence

As previously discussed in Section 3, the existence of self-similar relationships in network quantities at or near criticality was considered to be an indicator of continuous, second-order phase transitions. Researchers studying phase transitions caused by congestion also investigated the existence of self-similarity in relation to the onset of phase transitions, but for different variables than those studied by the percolation theory or epidemiologic researchers (see Table 2.2). Researchers using percolation-based approaches analytically observed self-similar relationships in quantities such as the growth of the giant connected component, the distribution of component sizes other than the giant component at criticality, and the distribution of distances between sites at criticality (see Sections 3 and 4). In contrast, congestion studies described in this section found self-similar relationships in quantities such as queue lengths, packet delay times, and load at sites. Their work suggested that self-similarity in these measured quantities may arise as a system approached a critical point. In addition, rather than determining power-law relationships analytically, empirical means were used in the congestion studies.

To identify self-similar processes, researchers used a variety of methods, such as computing the power spectrum over time series and producing log-log, or power-law, chart plots which reflected scale-invariance. The appendix discusses various methods for testing for the presence of self-similarity. Yet, the researchers involved were unable to derive a complete and consistent view of self-similar processes. In (Solé and Valverde, 2001), self-affinity in local fluctuations in the number of packets at λ_c was inferred using power spectral analysis of time series. This relationship was plotted on a log-log scale, where the slope was determined by $P(f) = f^{-\beta}$, $\beta = 0.97, \pm 0.06$. This study also found that the distribution of queue lengths "approached power laws" at λ_c. This study also observed an increase in variance of packet lifetimes as $\lambda \rightarrow \lambda_c$, followed by decrease in variance after $\lambda > \lambda_c$. In (Tadic, Rodgers, and Thurner, 2007), evidence for self-similar patterns in packet lifetimes and site queue lengths were detected through power spectral analysis at, and below, $\lambda < \lambda_c$. A log-log distribution in mean queue lengths, identified as a $1/f$ signal, sometimes referred to as "$1/f$ noise"[13], was observed at, and just below, λ_c (Mukherjee and Manna, 2005), as shown in Figure 8.

Self-similarity was found in the fluctuation of average load per site in (Mukherjee and Manna, 2005) just below the critical point. Other congestion studies that detected the presence of some degree of self-similarity included (Arrowsmith et al., 2004), where the phenomenon was detected in queue lengths below the critical point and found to rise as the system approached criticality. In (Woolf et al., 2004), self-similarity was observed in network traffic, as it was in other works. These findings experimentally demonstrated the existence of self-similarity in network models, but did not provide a complete picture of where and when self-similarity manifested itself.

[13] In the $1/f$ signal, f stands for frequency. The Appendix contains a discussion of the $1/f$ signal or $1/f$ "noise".

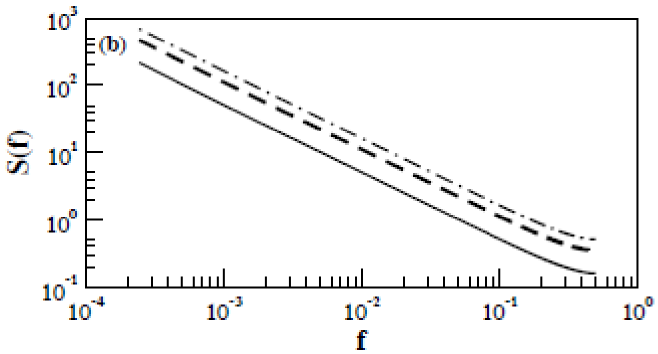


Figure 8. The Fourier transform of the autocorrelation function is computed on the mean queue length data to yield the power spectrum S(f), which is plotted against the frequency *f* on a double logarithmic scale for three values of λ, $\lambda = 0.96$ (dot-dashed), $\lambda = 0.97$ (dashed) and $\lambda = 0.98$ (solid) lines for a 64 x 64 lattice system, where λ is defined to be the probability at each time step that a new packet is injected into the network. The straightness of these curves indicates a power-law variation of the power spectrum: $S(f) \sim f^{-\phi(\lambda)}$, where $\phi(\lambda) \approx 1$, indicating the presence of the 1/*f* signal near $\lambda_c \approx 1$. The figure is from (Mukherjee and Manna, 2005).

Related to self-similarity is long-range dependence, which is generally understood to signify a pattern of dependence in time-series data which decays slowly, and hence persists over long time durations. In (Karagiannis, Molle, Faloutsos, 2004), long-range dependence occurs when "the behavior of a time-dependent process shows statistically significant correlations across large time scales". Like self-similarity, the presence–or absence–of long-range dependence in relation to the onset of phase transitions has been a topic of interest, and the results among researchers have proved somewhat inconclusive. To detect this phenomenon a variety of testing procedures were used (See the Appendix or details). Long-range dependence in time series was observed for router and host queue lengths at and below λ_c (Arrowsmith et al., 2004) using R/S statistical computations. In (Mukherjee and Manna, 2005), power spectral analysis showed that as the system approached λ_c, the fluctuation of the average load per site increased and the 1/*f* signal was detected, from which long-range dependence was inferred. In (Solé and Valverde, 2001) spectral analysis was also used to infer the existence of long-range dependence, though here the evidence appeared perhaps not as strong. In (Tadic, Rodgers, and Thurner, 2007), long-range correlations were detected below λ_c and found to decline as the system approached criticality. In (Woolf et al., 2002), the R/S statistic was used to find long-range dependence in distributions of packet delay times below and at criticality, but here the phenomenon appeared to be attributed to the "non-stationarity" of traffic sources. Long-range dependence was observed (perhaps not as strongly) near the phase transition point in a packet switching network (Lawniczak et al., 2007) and was also found in related studies of Internet traffic (Veres et al., 2003). In (Yuan and Mills, 2002), this phenomenon was observed in the behavior of a two-dimensional lattice network which employed different congestion control algorithms. Power spectra of the time series of packets transmitted to sites are used to detect the presence of the 1/*f* signal and to show long-range dependence increase in parts (subgraphs) of their model as model size increases and varies depending on traffic intensity and congestion control algorithm used. Overall, as in the case of self-similarity, long-range dependence was measured for isolated values of λ at, below, and above λ_c, rather than for a more complete ranges of load values. In (Yuan and Mills, 2002), long-range dependence was measured only when the network operated below criticality.

Despite the incompleteness, researchers studying congestion provided some characterization of self-similarity and long-range dependence in relation at, or near, the critical point. Yet, as with the study of the phase transition itself, much work needs to be done, because the full range of circumstances in which these phenomena may occur are also not well understood. Nor is it understood how these phenomena behave for wider ranges of control parameter values. Similarly, the origin of self-similarity is not well understood, as we will see shortly. Finally, it is notable that there has been almost no study of the *critical slowing down* phenomenon at the critical point, which was speculated about in (Solé et al., 1996) and briefly studied in (Sarkar et al., 2012). As mentioned previously (recall Section 1), in critical slowing down, a system that is approaching a continuous, second-order phase transition point is said to display an explicit pattern in which, in response to perturbations, measured behaviors increasingly deviate from equilibrium and recover more slowly. As the critical point is approached, the increase in critical slowing down can be measured and so serve as a warning of the impending phase transition. Researchers studying climate change and power grid blackouts have experimented with measurement methods for predicting phase transitions based on critical slowing down (Scheffler et al., 2009; Dakos et al., 2008; Hines, Cotilla-Sanchez, and Blumsack, 2011). Other than (Sarkar et al., 2012), some works did extensive work on self-similarity at, and near, the critical point, but did not identify critical slowing down, as for example (Fukuda, Takayasu, and Takayasu 2005). Therefore, critical slowing down in congestion studies of communication networks has not been fully explored. In particular, it would be desirable to know if critical slowing down can be related to self-similarity, which is believed to manifest itself prior to the onset of a continuous, second-order phase transition.

6.4 Related Studies of Self-Similarity in Models of Transmission Control Protocol Links

The incompleteness of our knowledge about self-similar behaviors in communication networks is underscored by the work of researchers who studied congestion in small network segments, which used the TCP/IP together with finite buffers and packet dropping. While the spatial scope of these models was very small, their efforts were significant because they show possible relationships between the occurrence of self-similarity and the operation of real-world protocols. Many works cited here discussed the origins of self-similarity but not all linked it to phase transitions and their onset.

In their experiments, these researchers increased the packet injection rate until, at a critical point, their systems underwent a transition to an overloaded state in which throughput severely declined and queues developed. They provided evidence of self-similarity by producing power-law and log-log graphs for such quantities as: distribution of congestion durations at the critical point (Fukuda, Takayasu, and Takayasu 2000; Fukuda, Takayasu, and Takayasu 2001; Takayasu, and Takayasu 2005) and in the variance of average throughput vs. duration for individual TCP/IP streams near and at criticality (Wisitpongphan and Peha, 2003). In (Fukuda, Takayasu, and Takayasu 2000; Fukuda, Takayasu, and Takayasu 2001; Fukuda, Takayasu, and Takayasu 2005), power spectrum analysis showed the $1/f$ signal pattern at the critical point, while away from the critical point, different power-law distributions emerged. In these studies, the source of self-similarity was attributed to TCP/IP protocol behaviors, such as the regularized transmission of acknowledgement packets in the TCP/IP feedback mechanism (Fukuda, Takayasu, and Takayasu 2001) and, to some degree, TCP/IP procedures generally (Fukuda, Takayasu, and Takayasu 2001; Wisitpongphan and Peha, 2003; Fukuda, Takayasu, and Takayasu, 2005).

In (Peha, 1997), the appearance self-similarity was also attributed to TCP/IP congestion control procedures under different load levels. This was also the case in the in-depth study reported in (Veres et al., 2003), who discussed how TCP/IP *propagates* self-similarity, although in these studies, phase transitions to an overloaded state were not simulated. In (Fukuda, Takayasu, and Takayasu, 2000), self-similarity was also attributed to a back off algorithm used in a simple model of Internet traffic. The work of (Fukuda, Takayasu, and Takayasu, 2000) also incorporated analysis of traces from real-world network traffic. The existence of self-similarity in Internet transmissions was determined by (Leland et al., 1994) using R/S measurements of Internet traffic samples, and since was also documented in other works, such as (Taqqu, Teverovsky, and Willinger, 1997; Crovella and Bestavros, 1997; Fukuda, Takayasu, and Takayasu 2000). In the case of (Veres et al., 2003), evidence of long-range dependence was also found, where the preservation of this pattern was attributed to the TCP/IP protocol. Also, note the work of (Erramilli et al., 2002), which concluded that the TCP/IP feedback control appears to modify self-similarity in network traffic, but it "neither generates it nor eliminates it".

A paper by (Yuan and Mills, 2006) reported investigations of network dynamics in a lattice model of routers (with a second tier composed of sources and receivers of transmission) and used the wavelet method (Abry and Vietch, 1998) to detect spatial-temporal correlations in small, medium, and large-scale traffic dynamics. Here, it was found that both TCP/IP congestion control and the alternative TCP-Friendly Rate Control protocol are associated with correlations over a limited range of timescales. However, the influence of both protocols was complex and differed under different circumstances with respect to long-range dependence, from which self-similarity was concluded. Much of the emphasis of (Yuan and Mills, 2006) was on abstraction of the network and representation of network variables (such as variability in file sizes), which were found to influence correlation.

The work by (Peha, 1997; Fukuda, Takayasu, and Takayasu 2000; Fukuda, Takayasu, and Takayasu 2001; Erramilli et al., 2002; Wisitpongphan and Peha, 2003; Veres et al., 2003; Takayasu, and Takayasu 2005; Yuan and Mills, 2006) and others raise the question of the source of self-similarity and long-range dependence in a communication network: does it always originate from traffic or can it also originate from an underlying protocol? If indeed protocols do generate self-similarity, it would be interesting to know the scope of the phenomenon: is it localized to a small set of links or is it network wide, and which network processes are affected? Finally, it is necessary to determine if it is possible to distinguish between protocol-generated signals emanating from local sources on the one hand, and signals associated with network traffic patterns having wider scope on the other. Being able to do this is necessary to provide an accurate assessment of the state of the network and to accurately predict the phase transition point. [14]

6.5 Modeling if Traffic Has a Poisson Distribution or is Self-Similar: Effects on Phase Transitions

There has been significant discussion on the nature of user sources, whether they have a Poisson distribution with respect to characteristics such as packet arrivals, connection establishment (or session establishment), or whether these characteristics exhibit self-similarity and long-range dependence. It should be pointed out that none of the works surveyed here attributed user behavior as a cause of whether or not phase transitions occur, but these works were more apt to determine if phase transitions occurred. Whether traffic sources have a Poisson distribution, or whether they are self-similar and also long-range dependent, may influence *where* phase transitions occur–and therefore, how the prediction of phase transitions should be modeled.

An article by (Paxson and Floyd, 1995), after examination of real-world traces, was among the first to argue that packet arrival did not have a Poisson distribution in nature but exhibited "bursty"

[14] In fact, the source of self-similarity in the network is a debate that has not ended and remains an open question that is beyond the scope of this survey.

characteristics that may be better modeled as being self-similar or long-range dependent. For session arrival times and connection establishment, the use of the Poisson distribution was found to be valid. The self-similar nature of Internet traffic was also concluded in (Leland et al., 1994; Crovella and Bestavros, 1997), which in some cases was also found to be long-range dependent (Leland et al., 1994). The survey by (Erramilli et al., 2002) pointed out that Internet traffic was concluded to be self-similar (and a Poisson distribution was limited to modeling session establishment and duration); with the origin of self-similarity arising from the nature of user sources or perhaps other factors discussed above. A refinement was later obtained by (Karagiannis et al., 2004), who reported that at sub-second intervals, a Poisson distribution was observed; but at larger time scales, the conclusions of (Leland et al., 1994) were confirmed. Then came the paper by (Gupta, Mahanti, and Ribiero, 2009), which found that traffic had a Poisson distribution at intervals of 5 s or less and was long-range dependent at longer intervals. The work by (Gupta, Mahanti, and Ribiero, 2009) is distinguished by its study of newer sources than (Karagiannis et al., 2004), including YouTube and Skype. However, these papers did not study phase transitions.

Among papers that considered the nature of user sources and addressed phase transitions, the Poisson distribution was used to model links in (Fukada, Takayasu, and Takayasu, 2005). In (Woolf et al., 2002; Woolf et al., 2004), packet arrival was modeled in different trials to alternatively have a Poisson distribution or to be long-range dependent. Here it was found that when traffic sources were designed to be long-range dependent rather than having a Poisson distribution, queues built up differently within the network, λ_c was at a different location, and phase transitions were more indistinct. With regards to study of phase transitions, traffic with a Poisson distribution and traffic that was long-range dependent were also compared in (Arrowsmith et al., 2004; Fukada, Takayasu, and Takayasu, 2001) and to some extent in (Mukherjee and Manna, 2005). Yet, many works on phase transitions in networks did not identify sources as having a Poisson distribution or as being self-similar. The topic of the modeling of self-similarity and long-range dependence has been a subject of interest to the network research community in general, and should be of interest to modelers of phase transitions as well. Modeling user behavior must take into account the nature of user sources in order to make accurate predictions about phase transitions.

6.6 Summary

Researchers studying models of congestion produced a significant amount of empirical evidence to characterize the transition from a global free state to a jammed state in random graph models of communication networks. Their work also included substantial evidence for the existence of scale invariance, or self-similarity, as increasing congestion caused the network to approach the critical point—although their conclusions were only preliminary. However, the variables for which self-similarity was discovered were different than the variables studied by the percolation theory and epidemiologic researchers (see Table 2.2). Moreover, the work on self-similarity was incomplete, and important questions remain to be answered regarding the entire range of circumstances under which it occurs, its scope, whether its origin is always independent of the underlying protocol, the extent to which it can be attributed to the nature of user sources, and the exact relationship between the appearance of self-similarity and phase transitions. The same set of questions apply to long-range dependence and the relationship of this phenomenon to that of the phase transition.

Although being substantial in size and more realistic than models developed in the previously described approaches in some respects, there was still a lack of realism in many of the models used for congestion studies with respect to topology, assumptions relating to behavior of buffers, as well as representation of real-world routing procedures and congestion control processes. As is generally the case on work discussed in this survey, the models were uniformly of one topology type, in contrast to the topologically heterogeneous Internet. The exception to this conclusion lies in the previously mentioned papers that reproduce TCP/IP procedures and congestion control safeguards on individual links, but the topologies modeled were very small. For these reasons, it is difficult to use the findings from the congestion studies to make accurate conclusions about real-world networks, such as the Internet, and this point has previously been made by others (Alderson and Willinger, 2005). This paper will return to the subject of improving model realism in Section 7.

Also of significance is the extent to which the role of congestion as a causal mechanism was fully described. In (Willinger et al., 2002) the point was made that much of the work in which emergent phenomena in networks was modeled is evocative rather than explanatory, i.e., the models used were capable of reproducing the phenomena of interest, but not in revealing their underlying causes. These authors described a *Model Validation Framework* for checking whether a proposed Internet-related model was evocative or explanatory. While application of this framework to the works surveyed here is beyond the scope of this study, one can say that in these studies, the buildup of congestion leading to a phase transition was generally viewed from a perspective exterior to the model. This exterior perspective consisted of periodically observing (and recording) variable values such as the average lifetime of packets, the number of packets in transit, and queue size at sites. However, the process by which congestion advanced through the network was itself not examined, and so the causes of phase transitions were not fully explained.

Despite the lack of detailed explanation, some authors attempt partial explanations, which suggested that a percolation process was being observed. For instance, in (Solé and Valverde, 2001), a distribution of queue lengths was found (which is) “probably the result of the presence of spatial structures, which propagate as waves”. The paper by (Echenique, Gomez-Gardenes and Moreno, 2005) stated that congested conditions began at the hubs and spread from there, forming impermeable aggregations of congested sites as well as regions of uncongested sites through which traffic flowed. Packets could flow only through regions that are uncongested, or permeable, which was compared to the flow of a liquid through a porous material. Finally, as load on the simulated system is increased as the system reached criticality, the number and extent of the impermeable regions increased, and the system transitioned to a jammed state. Similarly, in (Wu, Wang and Yeung, 2008), the authors stated that as the phase transition point was approached, congestion first took control of the hubs in the networks, i.e., hubs first entered a congested state. They went on to state that congestion rapidly invaded the whole network via a cascade, which progresses through the sites with smaller and smaller degrees and smaller betweeness. In (Wang et al., 2009b), congestion was speculated to start at a small number of sites and spread in cascading fashion. In (de Martino et al., 2009), the spread of congestion was speculated about as follows: “the mechanism triggering the emergence of congestion is somewhat reminiscent of jamming or bootstrap percolation, where a (site) is occupied if the number of occupied neighbors exceeds a given threshold”. Work by (Sarker et al., 2009; Rykalova, Levitan, and Brower, 2010; Sarkar et al., 2012) also studied how congestion spread, but relied on terminology and concepts of phase transitions that occurred in the thermodynamic limit. However, these papers made simplifying assumptions about topology and network protocols.

These partial explanations suggest the appropriateness of further investigating percolation theory and cascading mechanisms as explanations for how congestion spreads and causes phase transitions in the communication network models. This is an important question to be examined, because the relationships between percolation theory, cascade studies, and congestion studies, at present appear to be open and unresolved questions. (Recall also the distinctions drawn between the phase transitions caused by congestion-related cascades and the percolation transition by (Moreno et al., 2003)). Determining if percolation or cascades (or both) occur in communication networks models used in the congestion studies will require that appropriate experiments be designed and carried out. The larger question, raised in the introduction in Section 1, is whether global phase transitions in networks in general can be explained by percolation or cascading mechanisms. Both questions are left as topics for future work, which this survey will return to below.

6.7 Related Congestion Studies Not Involving Global Phase Transitions

In this section, we review the work of researchers who have studied complex phenomena in communication networks, which are sometimes related to phase transitions. These studies are categorized separately, because their goal was not the investigation of phase transitions, but assessment of the impact on the network of events such as attacks. While the events described in these studies led to congestion buildups that changed the state of the network, or at least a large subset of it, the scope of the attack was usually limited so that it did not lead to complete performance degradation. Nevertheless, these papers have many features in common with congestion studies described above and so are listed here.

Work by (Yuan and Mills, 2005a) showed how distributed denial-of-service attacks in network models led to local congestion, which resulted in emergent behaviors at a macroscopic level. Here, the effect of targeted denial-of-service attacks on routers is studied, using a network topology model in which TCP/IP was used for packet delivery. The study showed local congestion buildups caused by attacks led to the emergence of macroscopic spatial-temporal traffic patterns within the network model. The study concluded that, in response to such attacks, self-organization patterns are stronger in larger networks over longer time frames. A method for predicting such attacks was provided. In (Yuan and Mills, 2005b), a metric that employed eigenvector analysis on an inter-site cross-correlation matrix of time-series fluctuation measurements was used to identify occurrences of traffic buildup. This method was used to show such congestion buildups can cause large-scale effects beyond the target router by disturbing traffic flows and changing spatial-temporal traffic patterns in other correlated routers. This work focused on studying global effects of congestion restricted to a subset of sites in the simulated network. The question was not considered as to whether or not congestion could be extended to cause a global phase transition in which the entire network became inoperative. Work by (Woolf et al., 2004) also studied congestion in models, in which traffic sources both had a Poisson distribution and were long-range dependent (as well as self-similar), and in which TCP/IP was used. Here it was found that long-range dependent sources caused severe local congestion among highly connected sites that were in close proximity to each other (recall Section 6.5). A distinguishing characteristic of this work was the generation of inter-site links that were based on observed patterns in the Internet. However, like the work of Yuan and Mills, phase transitions were also not explicitly studied.

Among vulnerability analyses of network protocol behaviors that observed spreading affects, two examples are presented. In (Wu et al., 2007), results of an analysis of the vulnerability of Internet routing to common types of failures is presented. Here, the focus was Internet AS structure and the effects of routing protocols. Using a model of AS structure, the effects of network topology and routing procedures were assessed. Among other results, the study finds 32 % of the AS provider topologies modeled were vulnerable to a single-link failure, and that choices of routing policy could result in sub-optimal selection of routes. However, again the analysis of failure was limited to local congestion buildups and could not be extended to global behaviors. Work by (Sriram et al., 2006) considered targeted attacks on BGP peering sessions that exploited vulnerabilities in BGP Route Flap Damping procedures. Here, the circumstances under which these attacks cause routing disruptions and widespread isolation of AS provider networks was examined, using models based on known characteristics of Internet AS topologies and BGP routing policies and procedures. The study reported that attacks could result in substantial numbers of route withdrawals and isolation of AS prefixes (addresses). However, the focus of this work was on exploring BGP vulnerability (of "Route Flap Damping", in particular) and its potential effects, rather than investigation of the phenomenon as a phase transition.

Additional examples could be presented, but for lack of space. In the studies reviewed in this section, the scope of the phenomena is deliberately limited in order to accurately portray the circumstances of real-world attacks, which generally are also limited. To do this, these researchers incorporated realistic protocol behaviors into their models, and in the case of (Yuan and Mills 2005a; Sriram et al., 2006; Wu et al., 2007), realistic topologies also. One suspects that many of these studies could have been recast as phase transition studies, if the authors had wanted to do so. However, the use of these realistic elements also raises the question of investigating global phase transitions in models that incorporate real-world protocols and topologies. This is discussed in the next section.

7. Discussion and Future Work: the Need for Realistic Models of Communications Networks

This section discusses that overall state of knowledge about phase transitions in communication networks and summarizes shortcomings of this work. The section goes on to discuss future work needed to address these shortcomings. Three broad areas are considered: (1) identifying, or developing, a theory that explains how phase transitions occur in random graphs models of communication networks; (2) developing models of networks that incorporate real-world characteristics; (3) investigating if and how global phase transitions might occur in realistic models of communication networks.

7.1 Summary of State of Knowledge

With the exception of congestion studies described in Section 6, the work of researchers using the other approaches–the percolation theory approach, the epidemiologic approach, and the theoretical cascade studies–used percolation theory as a basis and provided a potential foundation for explaining how global, network-wide phase transitions occurred in random graph models of communication networks. Using this foundation, they developed mathematical formulae for estimating important quantities related to the percolation phase transition, the most important of these being the percolation threshold, or critical point, which served as an indicator of the onset of the phase transition. They also developed mathematical formulae that estimated the rate of growth of the giant connected component, which described the magnitude of the event (see Table 2.1). Another important aspect of their work was general agreement respect to power-law relationships, self-similarity, and other phenomena. Researchers who used percolation theory or epidemiologic approaches confirmed analytical results by simulation.

The percolation theory researchers assumed a generic spreading agent. Researchers using the epidemiologically based approach studied percolation in the context of SIS and SIR disease spreading processes to show how specific kinds of agents, such as viruses, spread through networks and reached stationary epidemic states. Cascade studies based on the percolation model focused on a specific spreading agent: the cascade or avalanche. Other cascade studies took an empirical approach, in which they observed phase transitions to a network-wide inoperable state that was associated with the breakup of the giant connected component. In this, their work echoed cascading studies that caused outages in electrical grids (Carreras et al., 2002).

Unlike the other approaches, most researchers who studied phase transition caused by network congestion did not attribute the transition from a free to a jammed state to percolation. Similarly, most of these researchers did not offer an alternative explanation for how global phase transitions occurred in communication networks. However, their empirical investigations contributed detailed information describing phase transitions caused by congestion in communication networks. Detailed measurements were provided to show the effects of rising congestion, as reflected in various global variables. In addition, the investigations included the study of self-similarity and long-range dependence, but these studies focused on quantities other than those considered by researchers who studied percolation processes (recall Table 2.2). Further, the study of self-similarity and long-range dependence was far from complete, and this survey identifies areas that require further investigation.

Most work in all four approaches was constrained to specific kinds of random graph structures and, in some cases, two-dimensional lattices. In most studies, topologies based on Internet structure and Internet protocol behaviors were not used, though some congestion studies were based on more realistic models and supplemented by analysis of data from real-world systems. Generally, the results produced by studies involving percolation of random graph networks indicated that the phase transitions were continuous. However, in the case of cascade studies that used percolation as a basis, (Watts, 2002; Gleeson and Cahalane, 2007) both continuous and discontinuous phase transitions were observed. Several congestion studies provided empirical confirmation of the existence of continuous phase transitions within random graph and lattice models. However, as in the case of cascades and transitions to jammed states, some studies also suggested that discontinuous phase transitions may also occur. Except studies that were based on percolation theory, most work that reported on phase transition order involved qualified characterization of the results. As mentioned, (Sarkar et al., 2009; Rykalova, Levitan, and Brower, 2010; Sarkar et al., 2012) described the phase transition in a finite, congested system using terminology and concepts from the theory of phase transitions. Perhaps more significantly, the highly important topic of metrics for predicting phase transitions was not covered in depth in any studies. However, analyses that predicted thresholds and empirical work describing self-similarity may provide basis for such metrics.

7.2 Toward a Theory of How Catastrophic Events Occur in Communication Networks

Perhaps first and foremost, it is necessary to identify, or develop, a theory that explains how and why catastrophic events occur in communication networks. How is this to be done? We have seen that percolation theory provided a basis for two of the four approaches discussed here—the approaches based on percolation theory itself and the epidemiologic approach. It was to some extent used as a theoretical basis in the third approach—based on cascades within networks. However, it was not often attributed in the fourth and largest category of studies, i.e., the congestion studies. We have also seen in the congestion studies that load was used as a mechanism for propagating cascades that led to phase transitions.

A reasonable first step is to investigate the use of percolation theory for explaining phase transitions observed in the congestion studies discussed in Section 6.3. This would entail revisiting or creating random graph models and message transfer processes of the type used in the congestion-related studies and then re-executing experiments, in which a control parameter, load, is increased until the phase transition occurs. However, rather than merely measuring outward manifestations of congestion, it will be necessary to also introduce a parameter, in the form of the growth of the giant connected component, which can be measured as the models transition to a jammed state. Thus, it may be possible to more rigorously determine whether the transition from a free state to a jammed state is an observed percolation transition. Similarly, it may be possible to determine if this transition occurs as a result of a cascading process (noting that percolation and cascade processes may not be mutually exclusive and both may apply). Otherwise, if this determination cannot be made, then an alternative explanation would be required for how congestion causes phase transitions.

Should the outcome of such an experiment result in a theory or theories of how phase transitions occur in communication networks, the relevance of the result would be limited to random graph models of communication networks. An outcome even in this limited context would be of value (note that (Sarkar et al., 2012) have already begun some of this work). Moreover, the result would potentially provide an important link to other scientific fields of endeavor, such as perhaps the study of phase transitions in power grids (Carreras et al., 2002; Newman et al., 2011), where the same causal explanations may be relevant. In fact, this raises the larger question of whether or not global phase transitions of the kind observed in random graph models of communication networks will occur in real-world networks, or in models of them. Answering this question leads us to a second direction for future research.

7.3 Making Models More Realistic

Realistic models of communication networks are necessary and realistic user behavior is necessary, if the results of researchers discussed in this paper are to be made relevant to real-world systems. Overall, researchers in all four groups utilized models which did not reflect real-world structures and processes, an observation also made by (Willinger et al., 2002; Alderson and Willinger, 2005; Dorogovtsev, Goltsev, and Mendes, 2008), among others. To this conclusion, there were exceptions, such as (Echenique, Gomez-Gardenes and Moreno, 2005), who used real-world network topologies as a basis in their models. There were also studies that accurately modeled TCP, including (Yuan and Mills, 2002; Yuan and Mills, 2005a) together the researchers studying TCP/IP flows over small topologies (Wisitpongphan and Peha, 2003; Fukuda, Takayasu, and Takayasu, 2005). However, lack of realism was generally the case, even when models of very substantial size were used. The reasons for this lack of reality in communication network models lay in three main areas: (1) lack of realistic representation of Internet-related protocols and procedures; (2) lack of realistic topologies in the models that the studies were based on; and (3) user behavior and its consequences. These three main areas each have separate subsections below.

7.3.1 Realistic representation of Internet-related protocols and procedures

One reason for the lack of reality in simulated models reviewed in this paper is that many studies (though not all) did not account for the congestion control mechanisms provided in TCP/IP (Mills et al., 2010), which mitigates congestion buildups. TCP/IP contains a feedback mechanism in the form of acknowledgement packets sent by destination sites to hosts which serve to regulate host transmissions and thus influence congestion patterns in the network. The congestion avoidance procedures of many variants of the TCP/IP protocol were extensively described and simulated in realistic Internet topologies by (Mills et al., 2010). Models that attempt to accurately portray Internet communications must carefully consider TCP/IP congestion control. Accurate models of TCP/IP must also include realistic models of finite buffers and, in the event of overflow, packet dropping procedures, and connection establishment procedures. Although a number of researchers included these considerations in their models that studied congestion in networks, overall, most studies did not.

Another reason for the lack of reality is related to the operation of routing protocols. Many researchers reviewed in this paper assumed either random choice of route (or a variable that selects the best route), shortest-path routing, or some combination based on these factors together with known traffic conditions and congestion levels. In fact, at the AS level routing decisions are often made on the basis of business agreements (which arguably could be modeled by random choice) and are also based on the routing procedures specified in BGP, which is widely used in the Internet. In addition, it is necessary to consider various contingency Internet routing behaviors in the face of attacks and failures (Sriram et al., 2006; Wu et al., 2007).

7.3.2 Realistic Topologies

This includes realism both in the sense of providing topological characteristics that are real and using a network model that has more than one type of topology (as for example the Abilene network). The lack of realistic topologies also reflects undifferentiated nodes in the model networks. For example in a real network, there are differences in packet sending rates between routers and host sites (which are generally leaf sites in a graph topology). There are also differences in sending rates among routers (which may differ considerably depending on router location) and among host sites. There are also propagation delays on links that cover distances. Also important is the lack of realistic assumptions about the packet arrival times at user sites and their effect on prediction (recall Section 6.5).

Most researchers reviewed in this survey chose to model Internet topology by using scale-free network models that were based on well-defined types of random graphs. In most cases, these scale-free networks were generated through a randomized process, which resulted in a large proportion of sites being linked to hub sites, through which most traffic flowed. Often these models were constructed using the Barabási-Albert preferential attachment algorithm (Barabási and Albert, 1999), described in Section 2. Moreover, the networks were modeled using a single type of topology, with no variation in logical structure across spatial extent. Similarly, in a real network, there are differences in packet sending rates at routers and leaf sites, where the latter represents users. Soon after the initial studies on scale-free networks, papers appeared that argued that the topology of the Internet, though scale-free, was structured differently. Most notably, in (Willinger et al., 2002; Alderson and Willinger, 2005), it was argued that Internet topology can be considered at two broad logical levels: (1) the global autonomous systems (AS) level topology, referred to here as the *Inter-AS level*, which consisted of AS that were interconnected in tiered hierarchies and in which each AS was a communication services provider that served a population of customers, and (2) the more local *Intra-AS level*, which referred to the individual network under the control of the AS and which is used by the AS to serve its customers.

The existence of AS-based hierarchies serves as an example of the topologically heterogeneity of the Internet (i.e., which is composed of multiple topology types), which contrasts with scale-free network models based on random graph topologies that assume a single type of topology. The definition of measurements described below, such as average degree and degree correlation, provide a basis for identifying additional heterogeneous characteristics in the topology of communication networks, which may vary in different physical sub-regions of real-world networks or within different AS networks. Topological heterogeneity is reinforced by differences among ASes in the implementation of routing policies, congestion control procedures, access controls, and packet forwarding speeds. It is also reinforced by differentiating nodes as hosts and routers, where each category can be differentiated as well and each category can have different packet forwarding speeds, routing policies, and access controls. All of these factors taken together, as well as factors to be discussed below, lead to a far more complicated and diverse picture of real-world networks. This picture is far more diverse than provided by topologically uniform models of scale-free networks based on random graphs which were assumed by most researchers in this survey.

The Inter-AS level. Within the Inter-AS level, it has been possible to describe well-known classes of peering relationships among AS providers, which provide a basis for defining tiered hierarchies (Wu et al., 2007; Oliveira et al., 2008). Two classes of relationships are most common: (1) *customer-provider*, in which customers reimburse providers for transport services; (2) and *peer-to-peer*, in which two providers cooperate to benefit from using each other's services. These inter-AS relationships are formed by business agreements made on the basis of factors such as market competition, user demand, and engineering considerations. The resulting topologies, while having scale-free distributions, may vary from the topologies produced by the Barabási-Albert preferential attachment algorithm for random graphs (Willinger et al., 2002; Liu et al., 2008).

Other work has suggested that, in addition to customer-provider and peering relationships, there are additional topological relationships that need to be considered at the Inter-AS level. For example, work by (Zhou and Mondragón, 2004b) examined maps of Internet topology generated from real-world data and detected the presence of highly interconnected groups of AS sites, referred to as the *rich clubs*. Sites within rich clubs tended to be more connected to each other than to other sites, a feature absent from Barabási-Albert scale-free networks (Zhou and Mondragón, 2004b; Woolf et al; 2004). Rich clubs have been found in some subsequent studies of Internet AS topology (Zhou, Cox, and Petricek, 2007; Zhou, Zhang and Zhang, 2007; Mondragón 2008). However, Internet AS level topologies were found not to have rich clubs due to their high bandwidth and high traffic capacity characteristics, though the rich club phenomena were found in other types of networks (Colizza et al., 2006). Similarly, degree correlations in the Internet were studied by (Mahadevan et al., 2006; Piraveenan, Prokopenko, and Zomaya, 2009; and others)[15].

The structure at the inter-AS level hierarchies has been investigated by a number of researchers who have proposed automated methods for inferring AS topology map approximations that could potentially be used in simulations, among the most notable and widely cited being (Gao, 2001; Mao, et al., 2003; Dimitropoulos et al., 2009) and more recently (Mahadevan et al., 2007; Dimitropoulos et al., 2007; Winter, 2009). For more surveys of work on inferring AS hierarchies, see (Oliveira et al., 2008; Haddadi et al., 2008). In (Mahadevan et al., 2007) a method was described for generating AS and router topologies (with router/AS memberships annotated) from Internet topological data, which preserved real-world statistical characteristics such as average degree, maximum degree, the distribution of distances between pairs of sites (as measured in path length). Work such as (Mahadevan et al., 2007), while providing a method for determining important statistical properties of real-world topologies, must be supplemented by other techniques for obtaining the actual topologies involved. Other methods for deriving models of Internet structure were developed by (Zhou and Mondragón, 2004a; Serrano, Boguñá and Díaz-Guilera, 2006; Piraveenan, Prokopenko, and Zomaya, 2009). In (Liu et al., 2008), a preference attachment algorithm that selected AS links on the basis of the number of users for an AS (as opposed to number of links) generated inter-AS topologies that had similar statistical properties to a BGP router network consisting of over 17 000 sites.

Despite extensive efforts, the overall structure of the Internet remains difficult to discern (Oliveira et al., 2010; Clegg, Di Cairano-Gilfedder and Zhou, 2010). Nevertheless, the knowledge gained from these studies appears sufficient to inform researchers seeking to develop more realistic models, and there have been a number of prototypes developed that generate models of Internet structure based on perceived Internet topological properties.

The Intra-AS level. The second topology level, pertaining to portions of the network that are owned or controlled by one AS, has been better articulated. As discussed previously, the topology of these networks has also been shown to be different from the random networks studied by researchers reviewed here. For instance, in (Alderson and Willinger, 2005), it was pointed out intra-AS Internet topologies were human engineered using the *highly or heuristically optimized tolerance or trade-offs (HOT)* conceptual framework (Carlson and Doyle, 2002). As a result, in order to optimize network performance, hubs were pushed toward the periphery or leaves of the network, while the high-traffic, high-throughput core had a much more robust homogenously connected structure without hubs. Although the degree distribution in HOT networks statistically matched the models of scale-free communication networks, the topologies were far different. An example of an intra-AS network was

[15] Degree correlation is for the most part not accounted for in the phase transition studies surveyed in this paper, however there are notable exceptions (Vazquez and Y. Moreno, 2003; Boguñá, Pastor-Satorras and Vespignani, 2003; Joo and Lebowitz, 2004; Tadic, Rodgers, and Thurner, 2007; Gleeson, 2008; Huang et al., 2011).

provided by the Abilene network model (Crovella and Kolaczyk, 2003), which articulated the structure in which the core is homogenously connected and hubs existed on the periphery. The behavior of a multi-tiered intra-AS network model was extensively studied under different congestion control regimes in (Mills et al., 2010), though without focus on phase transitions. As in the case of Inter-AS topologies, there have been numerous efforts to describe intra-AS topologies, and these also were reviewed by (Haddadi et al., 2008).

7.3.3 User Behavior

An important aspect of realism is user behavior. Although perhaps less work has been done on user behavior than in other areas of communication network modeling, it remains important. Only a few of papers surveyed here incorporate user behavior to a significant extent, including (Woolf et al., 2002; Woolf et al., 2004; Arrowsmith et al., 2004; Karagiannis et al., 2004; Fukada, Takayasu, and Takayasu, 2005; Mukherjee and Manna, 2005; Gupta, Mahanti, and Ribiero, 2009), and these considered user behavior indirectly. Recall also Section 6.5, which considers whether traffic arrival in a network should be modeled using a Poisson distribution, or whether it should exhibit long-range dependence. User behavior in the preceding studies was important primarily in determining the precise position of the phase transition, not whether it took place or not. As such, user behavior was seen, and continues to be seen, as an important factor in predicting when a phase transition will occur.

User behavior consists of two parts: (1) the behavior of human users, including what they submit for transmission, when they submit it, and when they find the network unusable due to congestion; and (2) the behavior on automated programs, which increasingly are a component of the Internet, as they contribute to the traffic being sent. The context and volume of both human and automated portions of user behavior are changing as the Internet continues to evolve, a fact acknowledged in (Gupta, Mahanti, and Ribiero, 2009). Hence, the exclusion of user behavior from models predicting phase transitions makes those models more unrealistic. Some studies that have included user behavior into models (though not to study phase transitions) were (Morris and Tay, 2003; Qiu, Liu, and Cho, 2005; Yu et al., 2006; Tay et al., 2008; Kim et al., 2008; Mills, Schwartz, and Yuan, 2010; Mills et al., 2010), though this is not an exhaustive list.

7.4 Studying Catastrophic Behaviors Using Realistic Models of Communication Networks

Extending simulation models of distributed communication systems to include realistic topologies, routing protocols, and TCP/IP congestion control raises a number of questions. The most obvious question is the extent to which phase transition observed by researchers using the four approaches in random network models could be replicated in models based on networks that are based on Inter- and Intra-AS topologies and include protocol behaviors. Perhaps more importantly, one could pose the question as to *whether phase transitions will occur at all* in such realistic models. The question is significant because congestion control protocols contain mechanisms to prevent spread of agents, such as congestion, which lead to phase transitions, while router protocols are designed to facilitate correct addressing and control the flow of traffic. In addition, the behavior of realistic models will need to be further modified by including known procedures designed to discover and combat spread of other agents, such as viruses and failure cascades. The inclusion of such protocol behaviors and defensive mechanisms will result in a significantly more complicated model than the models used by most of the researchers discussed in this paper.

Approaches to constructing realistic models. First, one must consider that it is nearly impossible and unrealistic to model the entire Internet or even large portions of it. Hence, it is only reasonable to focus on small portions, such as intra-AS networks or perhaps small combinations of AS networks, perhaps obtained by generating Internet topologies as discussed above. Working in this scope, the question needs to be considered within the context of interconnected network subcomponents, each having different but well-defined graph topologies, some of which may be analyzable using currently known random graph models, but which may also require use of different graph topology types. Likely, realistic topologies may include multiple topology types that are expressed at different levels of abstraction. Modeling new and different types of graph topologies may be aided by being able to isolate subcomponents of larger networks and study them separately.

To determine if there are circumstances under which phase transitions occur in realistic networks, researchers may be able to leverage studies on the observed percolation of subgraphs (Vazquez et al., 2004; Corominas-Murta, 2010) and on observed percolation processes involving communities of interconnected sites within larger networks, known as cliques (Derényi, Palla, and Vicsek, 2005; Palla, Derényi, and Vicsek, 2007). Other relevant work on subgraphs includes the previously mentioned studies of the rich club phenomenon (Zhou, Zhang and Zhang, 2007; Opsahl et al., 2008) and investigation of clusters of neighboring sites that are strongly connected (Soffer and Vázquez, 2005). Study of such structures provides a means of analyzing the spread of properties across more localized (and sometimes isolated) sets of sites, which may be more heavily interconnected and prone to interacting more intensively with each other than the rest of the network. Along these lines, the previously described work of (Zou, Towsley, and Gong, 2007) focused on the epidemic spread of email worms over logical networks (subnetworks) defined by email address relationships, rather than the more global topologies determined by network connectivity (recall Section 4.2). In addition to topological considerations, models of spreading dynamics that more accurately captured real-world processes may have to be incorporated. For example, (Willinger et al., 2002) discussed the possibility of successfully developing mathematical models that accurately represented in Internet traffic patterns and provided a validation framework to test applicability.

Detailed Characterization of Phase Transitions. If there are circumstances under which phase transitions are observed in models with Internet-like structures, other questions follow. For instance, are these phase transitions explainable by percolation theory or cascades? If not, how are they explainable? Can the phase transition be classified as a continuous or discontinuous phase transitions? Do both continuous and discontinuous phase transitions occur depending on the circumstances, as suggested in (Watts, 2002; Gleeson and Cahalane, 2007; Echenique, Gomez-Gardenes, and Moreno, 2005; Wu, Wang and Yeung, 2008; Buldyrev et al., 2010). The question is interesting, because if discontinuous, first-order phase transitions to states in which network performance deteriorates are possible in models with realistic characteristics, this may have potential consequences for reliability of real-world systems. The answers to these questions will require examination of phase transition order using models that portray a variety of real-world topologies and operational situations.

Self-Similarity, Long-Range Dependence, and the 1/f Signal. Another important area of future work is the creation of more detailed and complete characterizations of self-similarity, long-range dependence, and the 1/*f* signal in communication networks. Together with a deeper understanding, it is necessary to obtain a more precise understanding of the relationship of these phenomena and the onset of phase transitions. Knowledge about self-similarity and long-range dependence has thus far been largely limited to what has been learned either from abstract, topologically homogenous models based on random networks that incompletely reflect Internet topologies and protocol behaviors or from simple single-link models of TCP/IP transmission. These studies need to be extended to experimental simulations using network models that are both sizable and also include a variety of topologies that incorporate real-world elements discussed above. The studies will also have to be extended to examine the behavior of self-similarity and long-range dependence in different quantities at wide ranges of control variable values below and above the critical point. Comprehensive knowledge is lacking of the presence of self-similar behaviors at, and away from, the critical point. For instance, the presence of the $1/f$ signal (a form of self-similarity) has been reported in several papers discussed in this survey, including (Csabai, 1994; Yuan and Mills, 2002; Mukharjee and Manna, 2005; Fukuda, Takayasu, and Takayasu, 2005). However, to date there is no complete analysis of which measureable quantities within communication networks exhibit this phenomenon, or how the $1/f$ signal behaves in relation to the critical point (i.e., is it present or absent below, at, and above a critical point?). Thus, the study of the $1/f$ signal, and its relationship to phase transitions in communication networks, remains an important topic of future work.

Experiments will also have to be conducted to determine the source of signals that produce self-similarity and long-range dependence in order to learn if they are caused by characteristics of network traffic and if so, to identify the specific causes, or if self-similarity can be caused by the workings of network protocols. Likely, self-similarity and long-range dependence may have multiple causes, and if so, this must be determined. Here, it will also be possible and necessary to study this phenomenon in data produced directly by real-world networks, as researchers have already done in some cases. Overall, the goal of future work in this area should be to provide a more complete picture of self-similarity and long-range dependence, together with the relationship, if any, of these phenomena to phase transitions. A complete picture does not exist today.

Developing Methods to Predict Phase Transitions. Related to self-similarity is the previously discussed phenomenon of critical slowing down for systems approaching a continuous phase transition (recall Section 6.2). To date, there has been no study of critical slowing down in communication networks. Yet, this is a highly important area of work, because it will provide a basis to develop metrics that will enable the prediction of phase transitions in real-world systems. As mentioned above, work in this problem has begun in other scientific fields, e.g., to predict events such as climate shifts (Scheffer et al., 2009; Dakos et al., 2008) and blackouts in power grids (Hines, Cotilla-Sanchez, and Blumsack, 2011). The phenomenon receives some attention in (Fukuda, Takayasu, and Takayasu 2005), though it is not the main point of that paper. To undertake this work for communication networks, it will be desirable to leverage what work has been done on this problem in other domains. To properly construct such metrics for communication networks, it will be necessary to understand the behavior of critical slowing down under a wide variety of circumstances, and also to relate this phenomenon to observed self-similarity at the critical point. It is likely that the work on self-similarity done in the congestion studies as well as work on estimating thresholds done by the percolation theory and epidemiologic researchers provides a basis for developing metrics that predict phase transitions. Finally, phenomenon such as protocol generated self-similarity must be considered (recall Section 6.3) in order to be able to recognize and filter out these effects, to allow accurate measurements and predictions to be made.

8. Conclusions

This survey has discussed research on phase transitions, which lead to catastrophic events in communication networks. This discussion of this body of work has been organized along the lines of four distinct approaches which describe separable research communities. The survey has shown these four approaches differ with respect to their focus and methods of study, but that each has made important contributions to our knowledge of phase transitions in networks. These contributions include advancing theoretical understanding about how phase transitions occur, by describing observed properties of the phase transitions, or by providing insight into how specific agents cause phase transitions. The survey has also shown the existence of common themes among these groups, as well as important differences. Essential contributions made by each group include the establishment of bounds for key quantities, agreement with respect to power-law relationships, self-similarity, as well as other phenomena.

One common theme relates to a theory that explains how phase transitions occur in communication networks. The survey has shown that in three of these approaches, researchers look to percolation theory to provide such an explanation: Section 3 described the work of researchers who used percolation theory explicitly and directly; Section 4 described the work of researchers who combined percolation with models of disease spread; while in Section 5, some researchers explained cascades in terms of percolation theory (though others also employed empirical observation of network simulation to study cascade phenomena). A fourth approach, taken by most of the researchers who study phase transitions, studied the buildup of congestion due to excessive load. While, in the fourth approach, percolation was not explicitly been attributed as the cause of phase transitions due to congestion by most researchers, the possibility is worthy of investigation. This paper has argued that an appropriate line of inquiry is to determine whether or not global phase transitions in communication networks due to congestion is essentially a variant of a percolation transition. Related to this question is the need to more thoroughly investigate different phenomena, such the manifestation of self-similarity and long-range dependence in communication networks. Research has shown that these phenomena both appear and disappear in various measureable quantities as communication networks are forced to a critical point. There is a need to more thoroughly study these phenomena to understand how they manifest themselves in networks and their relationship to phase transitions. The knowledge gained from understanding these phenomena may provide a basis to also understand critical slowing down. This will further the understanding of, and development of, methods to predict the onset of phase transitions.

To develop methods to predict phase transitions will require that researchers overcome the lack of realism in the models used to study phase transitions in communication networks. The lack of realism in the structure and dynamics of real-world networks on the one hand and random graphs employed by researchers studying phase transitions on the other hand has been pointed out in earlier studies, and this survey agrees with this point. However, the use of abstracted representations of a problem should be viewed as an expected feature when work begins in a new research area. Investigating the theories for phase transitions and related phenomena must be done in realistic models to have validity for real-world networks. Accordingly, the survey has provided directions for future research intended to overcome this issue, most notably in the discussion on developing techniques for discovering and incorporating characteristics of real-world communication networks into simulation models. In the future, this work will ultimately be extended to develop both theoretic models and realistic simulations that accurately depict the dynamics of phase transitions in real-world networks. The ultimate goal of this work must be to develop predictive metrics, which will allow us to anticipate the onset of unfavorable macroscopic changes in networks that lead to catastrophic events, so that their widespread effects can be mitigated and even avoided.

9. References

P. Abry and D. Veitch, 1998. Wavelet Analysis of Long-Range Dependent Traffic*. IEEE Transactions on Information Theory*. Volume 44, Issue 1, pp. 2−15.

R. Albert, H. Jeong, and A. Barabási, 2000. Attack and error tolerance of complex networks. *Nature*. Volume 406, pp. 378–382.

D. Alderson and W. Willinger, 2005. A Contrasting Look at Self-Organization in the Internet and Next-Generation Communication Networks. *IEEE Communications Magazine*. Volume 43, Issue 7, pp. 94–100.

P. Allegrini, D. Menicucci, R. Bedini, L. Fronzoni, A. Gemignani, P. Grigolini, B. West, and P. Paradisi, Bo2009. Spontaneous brain activity as a source of ideal 1/f noise. *Physical Review E*. Volume 80, pp. 061914-1 – 061914-13.

D. Arrowsmith, R. Mondragón, J. Pitts, and M. Woolf, 2004. *Phase Transitions in Packet Traffic on Regular Networks a comparison of source types and topologies*. Report 08, 2004/2005. Institut Mittag-Leffler.

F. Ball and T. Britton, 2005. An Epidemic Model with Exposure-Dependent Severities. *Journal of Applied Probability*. Volume 42, pp. 932−945.

P. Bak, C. Tang, and K. Wiesenfeld, 1988. Self-Organized Criticality. *Physical Review A*. Volume 38, Number 1, pp. 364–375.

P. Bak, C. Tang, and K. Wiesenfeld, 1987. Self-Organized Criticality: An Explanation of 1/f Noise. *Physical Review of Letters*. Volume 59, Number 4, pp. 381–384.

A. Barabási and R. Albert, 1999. Emergence of Scaling in Random Networks. *Science*. Volume 286, pp. 509−512.

D. ben-Avraham and S. Havlin, 2000. *Diffusion and Reaction in Fractals and Disordered Systems*. Cambridge University Press, Cambridge England.

N. Berger, C. Borgs, J. Chayes, and A. Saberi, 2005. On the Spread of Viruses on the Internet. *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithm (SODA '05)*. Vancouver, Canada. pp. 301−310.

S. Boccalettia, V. Latora, Y. Moreno, M. Chavez, and D. Hwanga, 2006. Complex networks: Structure and dynamics. *Physics Reports*. Volume 424, pp. 175–308.

M. Boguñá, R. Pastor-Satorras, and A. Vespignani, 2003. Absence of Epidemic Threshold in Scale-Free Networks with Degree Correlations. *Physical Review Letters*. Volume 90, pp. 028701-1−028701-4.

B. Bollobás, 1984. The Evolution of Random Graphs*. Transactions of the American Mathematical Society. Volume 286, Number 1, pp. 257−274.*

B. Bollobás and F. Chung, 1988. The Diameter of a Cycle Plus a Random Matching. *SIAM Journal of Discrete Mathematics*. Volume 1, Issue 3, pp. 328−333.

B. Bollobás S. Janson, and O. Riordan, 2006. The Phase Transition in Inhomogeneous Random Graphs. *Random Structures and Algorithms. Volume* 31, Issue 1, pp. 3−122.

B. Bollobás and O. Riordan, 2006. *Percolation*. Cambridge University Press.

C. Borgs, J. Chayes, A. Ganesh, and A. Saberi, 2010. How to distribute an antidote to control epidemics. *Random Structures and Algorithms*. Volume 37, Issue 2, pp. 204−222.

M. Bramson, R. Durrett, and R. Schonmann, 1991. The Contact Process in a Random Environment. *The Annals of Probability*. Volume 19, Number 3, pp. 960−983.

S. Buldyrev, R. Parshani, G. Paul, H. Stanley, and S. Havlin, 2010. Catastrophic cascade of failures in interdependent networks. *Nature*. Volume 464, pp. 1025–1028.

D. Callaway, M. Newman, S. Strogatz, and D. Watts, 2000. Network Robustness and Fragility: Percolation on Random Graphs. *Physical Review E*. Volume 85, pp. 5468–5471.

J. Carlson and J. Doyle, 2002. Complexity and robustness. *Proceedings of the National Academy of Sciences*. Volume 99, supplement 1, pp. 2539–2545.

B. Carreras, V. Lynch, I. Dobson, and D. Newman, 2002. Critical points and transitions in an electric power transmission model for cascading failure blackouts. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. Volume 12, Number 4, pp. 985−994.

K. Christensen, 2002. *Percolation Theory*. Massachusetts Institute of Technology. http://www.mit.edu/~levitov/8.334/notes/percol_notes.pdf, 40 pages.

R. Clegg, 2006. A Practical Guide to Measuring the Hurst Parameter. *International Journal of Simulation: Systems, Science & Technology*. Volume 7, Number 2, pp. 3–14.

R. Clegg, C. Di Cairano-Gilfedder, and S. Zhou, 2010. A critical look at power law modelling of the Internet. *Computer Communications*. Volume 33, Issue 3, pp. 259–268.

E. Coffman, Z. Ge, V. Mishra, and D. Towsley, 2002. Network Resilience: Exploring Cascading Failures within BGP. *Proceedings of the 40th Annual Allerton Conference on Communications, Computing, and Control*. Monticello (Illinois), United States. pp. 97–106.

R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, 2000. Resilience of the Internet to Random Breakdowns. *Physical Review of Letters*. Volume 85, Number 21, pp. 4626–4628.

R. Cohen, K. Erez, D. Avaraham, and S. Havlin, 2001. Breakdown of the Internet under intentional attack. *Physical Review of Letters*. Volume 86, pp. 3682–3685.

R. Cohen, D. ben-Avraham, and S. Havlin, 2002. Percolation Critical Exponents in Scale-Free Networks. *Physical Review E*. Volume 66, pp. 036113-1 –036113-4.

V. Colizza, A. Flammini, M. Serrano, and A. Vespignani, 2006. Detecting rich-club ordering in complex networks. *Nature Physics*. Volume 2, pp. 110–115.

C. Cooper and A. Frieze, 2003. A general model for web graphs. *Random Structures & Algorithms*. Volume 22, Issue 3, pp. 311–335.

B. Corominas-Murtra, 2010. Percolation of arbitrary uncorrelated nested subgraphs. *The European Physical Journal B.* Volume B 77, Issue 2, pp. 213–217.

M. Crovella and A. Bestavros, 1997. Self-Similarity in World Wide Web Traffic—Evidence and Possible Causes. *IEEE/ACM Transactions on Networking*. Volume 5, pp. 835–846.

M. Crovella and E. Kolaczyk, 2003. Graph Wavelets for Spatial Traffic Analysis. *Proceedings of the IEEE INFOCOM*. Volume 3. San Francisco, United States. pp. 1848–1857.

I. Csabai, 1994. 1/f noise in computer network traffic. *Journal of Physics A: Mathematical and General*. Volume 27, Number 12, pp. L417–L421.

L. da Fontoura Costa, O. Oliveira, G. Travieso, F. Aparecido Rodrigues, P. Ribeiro Villas Boas, L. Antiqueira, M. Palhares Viana, L. Correa da Rocha, 2011. Analyzing and Modeling Real-World Phenomena with Complex Networks: A Survey of Applications. *Advances in Physics*. Volume 60, Issue 3, pp. 329–412.

V. Dakos, M. Scheffer, E. van Nes, V. Brovkin, V. Petoukhov, and H. Held, 2008. Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*. Volume 105, Number 38, pp. 14308–14312.

D. De Martino, L. Dall'Asta, G. Bianconi and M. Marsili, 2009. Congestion phenomena on complex networks. *Physical Review E.* Volume 79, pp. 015101-1−015101-4.

I. Derényi, G. Palla, and T. Vicsek, 2005. Clique percolation in random networks**.** *Physical Review Letters*. Volume 94, pp. 160202-1−160202-4.

C. Di Cairano-Gilfedder and R. Clegg, 2005. A decade of internet research: advances in models and practices. *BT Technology Journal*. Volume 23, Issue 4, pp. 115−128.

X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, K. Claffy, and G. Riley, 2007. AS Relationships: Inference and Validation. *ACM SIGCOMM Computer Communications Review.* Volume 37, Issue 1, pp. 29–40.

X. Dimitropoulos, D. Krioukov, A. Vahdat, and G. Riley, 2009. Graph Annotations in Modeling Complex Network Topologies. *ACM Transactions on Modeling and Computer Simulation*. Volume 19, Issue 4, pp. 17:1–17:29.

S. Dorogovtsev, A. Goltsev, and J. Mendes, 2008. Critical phenomena in complex networks. *Review of Modern Physics*. Volume 80, Issue 4, pp. 1275–1355.

R. Durrett, 2010. Some features of the spread of epidemics and information on a random graph. *Proceedings of the National Academy of Sciences*. Volume 107, Number 10, pp. 4491–4498.

D. Easley and J. Kleinberg, 2010. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press.

P. Echenique, J. Gomez-Gardenes and Y. Moreno, 2005. Dynamics of jamming transitions in complex networks. *Europhysics Letters* Volume 71, Number 2, pp. 056105-1– 056105-4.

P. Ehrenfast, 1933. Phasenumwandlungen im ueblichen und erweiterten Sinn, classifiziert nach dem entspechenden Singularitaeten des thermodynamischen Potentiales. *Verhandlingen der Koninklijke Akademie van Wetenschappen.* Volume 36, pp. 153–157; *Communications from the Physical Laboratory of the University of Leiden*, Supplement Number 75b (in German).

A. Erramilli, M. Roughan, D. Veitch, W. Willinger, 2002. Self-Similar Traffic and Network Dynamics. *Proceedings of the IEEE*. Volume 90, Number 5, pp. 800–819.

P. Erdös and A. Rényi, 1961. On the Evolution of Random Graphs. *Bulletin of the Institute of International Statistics*. Volume 38, Number 4, pp. 343–347.

K. Fukada, M. Takayasu, and H. Takayasu, 2000. Origin of critical behavior in Ethernet traffic. *Physica A: Statistical Mechanics and its Applications*. Volume 287, Issues 1–2, pp. 289–301.

K. Fukada, M. Takayasu, and H. Takayasu, 2001. Self-Similar Traffic Originating in the Transport Layer. *Communication Networks and Distributed Systems Modeling and Simulation Conference*. Phoenix, United States.

K. Fukuda, M. Takayasu, and H. Takayasu, 2005. A cause of self-similarity in TCP traffic. *International Journal of Communication Systems*. Volume 18, Issue 6, pp. 603–617.

A. Ganesh, L. Massoulié, and D. Towsley, 2005. The Effect of Network Topology on the Spread of Epidemics. *Proceedings of the IEEE INFOCOM 2005: 24th Annual Joint Conference of the IEEE Computer and Communications Societies (Volume 2).* Miami, United States. pp. 1455–1466.

L. Gao, 2001. On Inferring Autonomous System Relationships in the Internet. *IEEE/ACM Transactions on Networking*. Volume 9, Number 6, pp. 733–745.

C. Gershenson, 2003. *Phase Transitions in Random Boolean Networks with Different Updating Schemes*. arXiv:nlin/0311008v1.

J. Gleeson and D. Cahalane, 2007. Seed size strongly affects cascades on random networks**.** *Physical Review E*. Volume 75, pp. 056103-1−056103-4.

J. Gleeson, 2008. Cascades on correlated and modular random networks. *Physical Review E*. Volume 77, pp. 046117-1−046117-10.

P. Grassberger, 1983. On the Critical Behavior of the General Epidemic Process and Dynamical Percolation. *Mathematical Biosci*ences. Volume 63, Number 2, p. 157–172.

G. Grimmett, 1989. *Percolation*. Springer-Verlag.

H. Gupta, A. Mahanti, and V. Ribiero, 2009. Revisiting Coexistence of Poissinity and Self-Similarity in Internet Traffic. *IEEE International Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems*. London, United Kingdom. pp. 1–10.

A. Hackett, S. Melnik, and J. Gleeson, 2011. Cascades on a class of clustered random networks. *Physical Review E*. Volume 83, pp. 056107-1–056107-10.

H. Haddadi, M. Rio, G. Iannaccone, A. Moore, and R. Mortier, 2008. Network Topologies: Inference, Modeling, and Generation. *IEEE Communications Surveys & Tutorials*. Volume 10, Issue 2, pp. 48–69.

J. Halley, W. Kunin, 1999. Extinction risk and the 1/f family of noise models. *Theoretical Population Biology*. Volume 56, Number 3, pp. 215–230.

T. Harris, 1974. Contact Interactions on a Lattice. *The Annals of Probability*. Volume 2, Number 6, pp. 969–998.

G. Heinzel, A. Rüdiger and R. Schilling, 2002. *Spectrum and spectral density estimation by the Discrete Fouriertransform (DFT), including a comprehensive list of window functions and some new at-top windows*. Max Planck Institute for Gravitational Physics. http://hdl.handle.net/11858/00-001M-0000-0013-557A-5. 84 pages.

P. Hines, E. Cotilla-Sanchez, and S. Blumsack, 2011. Topological Models and Critical Slowing Down: Two Approaches to Power System Blackout Risk Analysis. *Proceedings of the 44th Hawaii Conference on System Sciences*. Kauai, United States. pp. 1–10.

H. Hinrichsen, 2006. Non-equilibrium phase transitions. *Physica A: Statistical Mechanics and its Applications*. Volume 369, Issue 1, pp. 1–28.

P. Holme and B. Kim, 2002. Vertex overload breakdown in evolving networks. *Physical Review E*. Volume 65, pp. 066109-1–066109-8.

E. Hooge, 1994. 1/*f* Noise Sources. *IEEE Transactions on Electronic Devices*. Volume 41, Number 11, pp. 1926–1935

Y. Hu, B. Ksherim, R. Cohen, and S. Havlin, 2011. Percolation in Interdependent and Interconnected Networks: Abrupt Change from Second to First Order Transition. *Physical Review E*. Volume 84, pp. 066116-1–066116-6.

Y. Hu, D. Zhou, R. Zhang, Z. Han, and S. Havlin, 2013. Percolation in interdependent networks with inter-similarity. *Physical Review E*. Volume 88, pp. 052805-1–052805-15.

X. Huang, J. Gao, S. Buldyrev, S. Havlin, S., and H. Stanley, 2011. Robustness of interdependent networks under targeted attack. *Physical Review E*. Volume 83, pp. 065101-1–065101-4.

G. Jaeger, 1998. The Ehrenfest Classification of Phase Transitions: Introduction and Evolution. *Archive for History of Exact Sciences*. Volume 53, Issue 1, pp. 51–81.

S. Janson, T. Luczak, T. Turova, and T. Vallier, 2012. Bootstrap Percolation on a Random Graph. *The Annals of Applied Probability*. Volume 22, Number 5, pp. 1989–2047.

J. Joo and J. Lebowitz, 2004. Behavior of susceptible-infected-susceptible epidemics on heterogeneous networks with saturation. *Physical Review E*. Volume 69, pp. 066105-1–066105-6.

L. Kadanoff, 2009. More is the Same; Phase Transitions and Mean Field Theories. *Journal of Statistical Physics*. Volume 137, Issue 5, pp. 777–797.

T. Karagiannis, M. Molle, and M. Faloutsos, 2004. Long-Range Dependence: Ten Years of Internet Traffic Modeling. *IEEE Internet Computing*. Volume 8, Number 5, pp. 57–64.

T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido, 2004. A Non-Stationary Poisson View of Internet Traffic. *Proceedings of INFOCOM 2004: Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies (Volume 3)*. Hong Kong, China. pp. 1558–1569.

W. Kermack and A. McKendrick, 1927. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. Volume 115, Issue 772, pp. 700–721.

H. Kim, K, Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Lee, 2008. Internet Traffic Classification Demystified: Myths, Caveats, and the Best Practices. *Proceedings of the ACM SIGCOMM Conference on Emerging Network Experiment and Technology (CoNEXT)*, Madrid, Spain, pp. 121–132.

Y. Lai, A. Motter, and T. Nishikawa, 2004. Attacks and Cascades in Complex Networks**.** *Lecture Notes in Physics*. Volume 650, pp. 299–310.

A. Lawniczak, P. Lio, S. Xie, and J. Xu, 2007. Study of Packet Traffic Fluctuations Near Phase Transition Point From Free Flow to Congestion in Data Network Model. *2007 Canadian Conference on Electrical and Computer Engineering: CCECE 2007.* Vancouver, Canada. pp. 360–363.

E. Lee, K. Goh, B. Kahng, and D Kim, 2005. Robustness of the avalanche dynamics in data packet transport on scale-free networks. *Physical Review E*. Volume 71, pp. 056108-1–056108-4.

W. Leland, M. Taqqu, W. Willinger, and D. Wilson, 1994. On the Self-Similar Nature of Ethernet Traffic. *IEEE/ACM Transactions on Networking*. Volume 2, Issue 1, pp. 1–15.

D. Levitin, P. Chordia, V. Menon, 2012. Musical rhythm spectra from Bach to Joplin obey a 1/f power law. *Proceedings of the National Academy of Sciences*. Volume 109, Number 10, pp. 3716–3720.

W. Li, A. Bashan, S. Buldyrev, H. Stanley, and S. Havlin, 2012. Cascading Failures in Interdependent Lattice Networks: The Critical Role of the Length of Dependency Links. *Physical Review Letters*. Volume 108, pp. 228702-1−228702-5.

T. Liggett, 1992. The Survival of One-Dimensional Contact Processes in Random Environments. *The Annals of Probability*. Volume 20, Number 2, pp. 696–723.

T. Liggett, 1999. *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes*. Springer-Verlag.

Y. Liu, J. Yuan, X. Shan, Y. Ren, and Z. Ma, 2008. Self-determined mechanisms in complex networks. *Physica A: Statistical Mechanics and its Applications*. Volume 387, Issue 8-9, pp. 2145–2154.

P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, 2006. Systematic Topology Analysis and Generation Using Degree Correlations. *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (SIGCOMM '06). Pisa, Italy. pp. 135–146.

P. Mahadevan, C. Hubble, D. Krioukov, B. Huffaker, and A. Vahdat, 2007. Orbis: Rescaling Degree Correlations to Generate Annotated Internet Topologies. *Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '07*). Kyoto, Japan. Volume 37, Issue 4, pp. 325–336.

Z. Mao, J. Rexford, J. Wang, and R. Katz, 2003. Towards an Accurate AS-Level Traceroute Tool. *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (SIGCOMM '03). Karlsruhe, Germany. pp. 365–378.

K. Mills, J. Filliben, D. Cho, E. Schwartz and D. Genin, 2010. *Study of Proposed Internet Congestion Control Mechanisms*, NIST Special Publication 500–282, 534 pages.

K. Mills, E. Schwartz, and J. Yuan, 2010. How to Model a TCP/IP Network Using Only 20 Parameters. *Proceedings of the 2010 Winter Simulation Conference*. Baltimore, MD. pp. 849–860.

E. Milotti, 2002. *1/f noise: a pedagogical review*. arXiv:physics/0204033, 26 pages.

M. Molloy and B. Reed, 1995. A Critical Point for Random Graphs with a Given Degree Sequence. *Random Structures and Algorithms*. Volume 6, Issue 2-3, pp. 161–180.

C. Moore and M. Newman, 2000. Exact solution on bond percolation in the small-world model. *Physics Review E.* Volume 62, pp. 7059–7064.

R. Mondragón, 2008. Topological modelling of large networks. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*. Volume 366, Number 1872, pp. 1931-1940.

Y. Moreno, R. Pastor-Satorras and A. Vespignani, 2002. Epidemic outbreaks in complex heterogeneous networks. *European Physical Journal B*. Volume 26, pp. 521–529.

Y. Moreno, J. Gomez, and A. Pacheco, 2002. Instability of scale-free networks under node-breaking avalanches. *Europhysics Letters*. Volume 58, Number 4, pp. 630–636.

Y. Moreno, R. Pastor-Satorras, A. Vazquez, and A. Vespignani, 2003. Critical load and congestion instabilities in scale-free networks. *Europhysics Letters*. Volume 62, Number 2, pp. 292–298.

R. Morris and Y. Tay, 2003. *A Model for Analyzing the Roles of Network and User Behavior in Congestion Control.* Technical report MIT–LCS–TR–898.

A. Motter and Y. Lai, 2002. Cascade-based attacks on complex networks. *Physical Review E*. Volume 66, pp. 065102-1−065102-4.

G. Mukherjee and S. Manna, 2005. Phase transition in a directed traffic flow network. *Physical Review E*. Volume 71, pp. 066108-1−066108-6.

H. Nakanishi and H. Stanley, 1978. A test of scaling near the bond percolation threshold? *Journal of Physics A: Mathematical and General*. Volume 11, Number 8, pp. L189–L197.

M. Newman, 2003. The Structure and Function of Complex Networks. *Siam Review.* Volume 45, Issue 2, pp. 167—256.

D. Newman, B. Carreras, V. Lynch, and I. Dobson, 2011. Exploring Complex Systems Aspects of Blackout Risk and Mitigation. *IEEE Transactions on Reliability*. Volume 60, Issue 1, pp. 134–143.

M. Newman, S. Strogatz, and D. Watts, 2001. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*. Volume 64, pp. 026118-1–026118-17.

B. Ninness, 1998. Estimation of 1/f Noise. *IEEE Transactions on Information Theory.* Volume 44, Issue 1, pp. 32–46.

T. Ohira and R. Sawatari, 1998. Phase Transition in Computer Network Traffic Model. *Physical Review E*. Volume 58, pp. 193–195.

R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. 2008. In Search of the Elusive Ground Truth: The Internet's AS-Level Connectivity Structure. ACM *SIGMETRICS Performance Evaluation Review.* Volume 36, Number 1, pp. 217–228.

R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, 2010. The (In)Completeness of the Observed Internet AS-Level Structure. *IEEE/ACM Transactions on Networking*. Volume 18, Issue 1, pp. 109–122.

T. Opsahl, V. Colizza, P. Panzarasa, and J. Ramasco, 2008. Prominence and Control: The Weighted Rich-Club Effect. *Physical Review Letters*. Volume 101, pp. 168702-1−168702-4.

G. Palla, I. Derényi, and T. Vicsek, 2007. The critical point of *k*-clique percolation in the Erdős–Rényi graph. *Journal of Statistical Physics*. Volume 128, Number1/2, pp. 219–227.

R. Pastor-Satorras and A. Vespignani, 2001a. Epidemic dynamics and endemic states in complex networks. *Physical Review E*. Volume 6, pp. 066117-1−066117-8.

R. Pastor-Satorras and A. Vespignani, 2001b. Epidemic Spreading in Scale-Free Networks. *Physical Review Lett*ers. Volume 86, pp. 3200–3203.

R. Pastor-Satorras and A. Vespignani, 2002a. Epidemic dynamics in finite size scale-free networks. *Physical Review E*. Volume 65, pp. 035108-1−035108-4.

R. Pastor-Satorras and A. Vespignani, 2002b. Immunization of complex networks. *Physical Review E*. Volume 65, pp. 036104-1–036104-8.

R. Pastor-Satorras and A. Vespignani, 2003. Epidemics and immunization in scale-free networks. In *Handbook of Graphs and Networks: From the Genome to the Internet*, S. Bornholdt and H. Schuster (eds.). Wiley Press. 22 pages.

V. Paxson and S. Floyd, 1995. Wide-Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking*. Volume 3, Issue 3, pp. 226–244.

J. Peha, 1997. Protocols Can Make Traffic Appear Self-Similar. *Proceedings of the 1997 IEEE/ACM/SCS Communication Networks and Distributed Systems Modeling and Simulation Conference*. San Diego, United States, pp. 47–52.

R. Pemantle, 1992. The Contact Process on Trees. *The Annals of Probability*. Volume 20, Number 3, pp. 2089–2116.

B. Pilgram and D. Kaplan, 1998. A comparison of estimators for 1/f noise. *Physica D: Nonlinear Phenomena.* Volume 114, Issues 1–2, pp. 108–122

M. Piraveenan, M. Prokopenko, and A. Zomaya, 2009. Local assortativity and growth of Internet. *European Physical Journal B*. Volume 70, pp. 275–285.

F. Qiu, Z. Liu, and J. Cho, 2005. Analysis of User Web Traffic with a Focus on Search Activities. P*roceedings of the Eighth International Workshop on the Web and Databases*. Baltimore, United States. pp. 103–108.

Y. Rykalova, L. Levitan, R. Brower, 2010. Critical phenomena in discrete-time interconnection networks. *Physica A: Statistical Mechanics and its Applications*. Volume 389, Issue 22, pp. 5259–5278

B. Samuelsson and J. Socolar, 2006. Exhaustive percolation on random networks. *Physical Review E*. Volume 74, pp. 036113-1–036113-17.

S. Sarkar, K. Mukherjee, A. Srivastav and A. Ray, 2009. Understanding phase transition in communication networks to enable robust and resilient control. *2009 American Control Conference (ACC '09)*, St. Louis, United States. pp. 1549–1554.

S. Sarkar, K. Mukherjee, A. Ray, A. Srivastav and T. Wettergren, 2012. Statistical Mechanics-Inspired Modeling of Heterogeneous Packet Transmission in Communication Networks. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*. Volume 42, Number 4, pp. 1083–1094.

G. Scalia-Tomba, 1985. Asymptotic Final-Size Distribution for Some Chain Binomial Processes. *Advances in Applied Probability*. Volume 17, Number 5, pp. 477–495.

M. Scheffer, J. Bascompte, W. Brock, V. Brovkin, S. Carpenter, V. Dakos, H. Held, E. van Nes, M. Rietkerk, and G. Sugihara, 2009. Early warning signals for critical transitions. *Nature*. Volume 461, pp. 53–59.

Y. Schwarzkopf, A. Rakos, and D. Mukamel, 2010. Epidemic spreading in evolving networks. *Physical Review E*. Volume 82, pp. 036112-1 – 036112-8.

M. Serrano, M. Boguñá and A. Díaz-Guilera, 2006. Modeling the Internet. *The European Physical Journal B – Condensed Matter and Complex Systems*. Volume 50, Issue 1, pp. 249–254.

R. Smith, 2011. The Dynamics of Internet Traffic: Self-Similarity, Self-Organization, and Complex Phenomena. *Advances in Complex Systems*. Volume 14, Issue 6, pp. 905–994.

S. Soffer and A. Vázquez, 2005. Network clustering coefficient without degree-correlation biases. *Physical Review E*. Volume 71, pp. 057101-1 – 057101-4.

R. Solé, S. Manrubia, B. Luque, J. Delgado, and J. Bascompte, 1996. Phase Transitions and Complex Systems: Simple, nonlinear models capture complex systems at the edge of chaos. *Complexity*. Volume 1, Issue 4, pp. 13–26.

R. Solé and S. Valverde, 2001. Information transfer and phase transitions in a model of internet traffic. *Physica A: Statistical Mechanics and its Applications*. Volume 289, pp. 595–605.

K. Sriram, D. Montgomery, O. Borchert, O. Kim, and D. Kuhn, 2006. Study of BGP Peering Session Attacks and Their Impacts on Routing Performance. *IEEE Journal on Selected Areas in Communications: Special issue on High-Speed Network Security*. Volume 24, Issue 10, pp. 1901–1915.

D. Stauffer and A. Aharony, 1994. *Introduction to Percolation Theory*. Taylor and Francis.

B. Tadic, G. Rodgers, and S. Thurner, 2007. Transport on Complex Networks: Flow, Jamming and Optimization. *International Journal of Bifurcation and Chaos*. Volume 17, Issue 7, pp. 2363–2385.

M. Takayasu, 1993. Characterization of violent fluctuations by interval distributions of level sets. *Physica A: Statistical Mechanics and its Applications*. Volume 197, Issue 3, pp. 371–378.

M. Taqqu, V. Teverovsky, and W. Willinger, 1997. Is Network Traffic Self-Similar or Multifractal? *Fractals.* Volume 5, pp. 63–73.

Y. Tay, D. Tran, E. Liu, W. Ooi, and R. Morris, 2008. Equilibrium analysis through separation of user and network behavior. *Computer Networks*. Volume 52, Issue 18, pp. 3405–3420.

USC ISI: University of Southern California, Information Sciences Institute, 2011. *DHS to Underwrite 5-Year, $16 Million Advance of ISI's Cyber Security Testbed.* http://www3.isi.edu/about-news_story.htm?s=240, Accessed September 4, 2015.

P. Van Mieghem, J. Omic, and R. Kooij, 2009. Virus Spread in Networks. *IEEE/ACM Transactions on Networking*. Volume 17, Issue 1, pp. 1–14.

A. Vazquez and Y. Moreno, 2003. Resilience to damage of graphs with degree correlations. *Physical Review E*. Volume 67, pp. 015101-1– 015101-4.

A. Vazquez, R. Dobrin, D. Sergi, J. Eckmann, Z. Oltvai, and A. Barabási, 2004. The topological relationship between the large-scale attributes and local interaction patterns of complex networks. *Proceedings of the National Academy of Sciences*. Volume 101, Number 52, pp. 17940–17945.

A. Veres, Z. Kenesi, S. Molnár, and G. Vattay, 2003. TCP's Role in the Propagation Of Self-Similarity in the Internet. *Computer Communications*. Volume 26, Number 8, pp. 899–913.

A. Vespignani and S. Zapperi, 1998. How self-organized criticality works: A unified mean-field picture. *Physical Review E*. Volume 57, pp. 6345–6362.

Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, 2003. Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint. *Proceedings of the 22nd International Symposium on Reliable Distributed Systems*. Pittsburgh, United States. pp. 25–43.

D. Wang, N. Cai, Y. Jing, and S. Zhang, 2009a. Phase Transition in Complex Networks. *American Control Conference: (ACC '09)*. St. Louis, United States. pp. 3310–3313.

W. Wang, Z. Wu, R. Jiang, G. Chen, and Y. Lai, 2009b. Abrupt transition to complete congestion on complex networks and control. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. Volume 19, pp. 033106-1–033106-7.

L. Ward and P. Greenwood, 2007. 1/f noise. *Scholarpedia*. Volume 2, Number 12, p. 1537.

D. Watts and S. Strogatz, 1998. Collective dynamics of 'small-world' networks. *Nature*. Volume 393, pp. 440–442.

D. Watts, 2002. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences of the United States of America*. Volume 99, Number 9, pp. 5766–5771.

W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker, 2002. Scaling phenomena in the Internet: Critically examining criticality. *Proceedings of the National Academy of Sciences of the United States of America*. Volume 99, Supplement 1, pp. 2573–2580.

R. Winter, 2009. Modeling the Internet Routing Topology - in less than 24h. *Proceedings of the 2009 ACM/IEEE/SCS 23rd Workshop on Principles of Advanced and Distributed Simulation* (PADS '09). Lake Placid, United States, pp. 72–79.

N. Wisitpongphan and J. Peha, 2003. Effect of TCP on Self-Similarity of Network Traffic. *Proceedings of 12th IEEE International Conference on Computer Communications and Networks (ICCCN)*. Dallas, United States, pp. 370–373.

M. Woolf, D. Arrowsmith, R. Mondragón, and J. Pitts, 2002. Optimization and phase transitions in a chaotic model of data traffic. *Physical Review E*. Volume 66, pp. 046106-1–046106-11.

M. Woolf, D. Arrowsmith, R. Mondragón, J. Pitts, and S. Zhou, 2004. Dynamical modeling of TCP packet traffic on scale-free networks. *Institut Mittag-Leffler*. Volume 6. 9 pages. http://www.researchgate.net/publication/245986610

J. Wu, Y. Anang, A. Mao, and K. Shin, 2007. Internet Routing Resilience to Failures: Analysis and Implications. *Proceedings of the 2007 ACM CoNEXT conference (CoNEXG '07).* New York, United States, pp. 25:1–25:12.

Z. Wu, W. Wang and K. Yeung, 2008. Traffic dynamics in scale-free networks with limited buffers and decongestion strategy. *New Journal of Physics.* Volume 10, pp. 023025-1 – 023025-12.

H. Yu, D. Zheng, B. Zhao, and W. Zheng, 2006. Understanding User Behavior in Large-Scale Video-on-Demand Systems. *Proceedings of the 1st ACM SIGOPS/EuroSys European Conference on Computer Systems*. Leuven, Belgium. Volume 40, Issue 4, pp. 333–344.

J. Yuan and K. Mills, 2002. Exploring Collective Dynamics in Communication Networks. *Journal of Research of the National Institute of Standards and Technology*. Volume 107, Number 2, pp. 179–191.

J. Yuan and K. Mills, 2005a. Monitoring the Macroscopic Effect of DDoS Flooding Attacks. *IEEE Transactions on Dependable and Secure Computing*. Volume 2, Issue 4, pp. 324–335.

J. Yuan and K. Mills, 2005b. A Cross-Correlation-Based Method For Spatial Temporal Traffic Analysis. *Performance Evaluation.* Volume 61, Issue 2, pp. 163–180.

J. Yuan and K. Mills, 2006. Simulating Timescale Dynamics of Network Traffic Using Homogeneous Modeling. *Journal of Research of the National Institute of Standards and Technology*. Volume 111, Number 3, pp. 227–242.

L. Zhao, K. Park, and Y. Lai, 2004. Attack vulnerability of scale-free networks due to cascading breakdown. *Physical Review E*. Volume 70, pp. 035101-1–035101-4.

J. Zhao, H. Zhou, and Y. Liu, 2013. Inducing Effect on the Percolation Transition in Complex Networks. *Nature Communications*. Volume 4, Number 12, pp. 2412-1–2412-5.

S. Zhou and R. Mondragón, 2004a. Accurately modeling the Internet topology. *Physical Review E*. Volume 70, Number 6, pp. 066108-1– 066108-8.

S. Zhou and R. Mondragón, 2004b. The Rich-Club Phenomenon in the Internet Topology. *IEEE Communications Letters*. Volume 8, Number 3, pp. 180–182.

S. Zhou, I. Cox, and V. Petricek, 2007. Characterising Web Site Link Structure. *Ninth IEEE International Workshop on Web Site Evolution (WSE 2007)*. Paris, France, pp. 73–80.

S. Zhou, G. Zhang, and G. Zhang, 2007. Chinese Internet AS-Level Topology. *IET Communications*. Volume 1, Number 2, pp. 209–214.

C. Zou, D. Towsley, and W. Gong, 2007. Modeling and Simulation Study of the Propagation and Defense of Internet Email Worm. *IEEE Transactions on Dependable and Secure Computing*. Volume 4, Issue 2, pp. 105–118.

Appendix: Definitions of Self-Similarity and Long-Range Dependence

In (Karagiannis, Molle, and Faloutsos, 2004), self-similarity is intuitively described as "the phenomenon in which the behavior of a process is preserved irrespective of scaling in space or time". Self-similarity is also defined in Section 1.2. As indicated previously (recall Section 6.3), long-range dependence means that the "behavior of a time-dependent process shows statistically significant correlations across large time scales" (Karagiannis, Molle, Faloutsos, 2004). For the interested reader, this appendix provides additional information on how these phenomena were tested for and discovered in the works covered by this survey, together with references that provide additional details for each measurement method. Much of the information provided here is reproduced from (Smith, 2011; Clegg, Di Cairano-Gilfedder, and Zhou, 2010) for the reader's convenience.

At an abstract level, in a stationary, continuous time process, $X(t)$, $t \geq 0$, the existence of self-similarity can be characterized by the expression:

$$X(t) = c^{-H} X(ct) \qquad \text{(A1)}$$

where c is a constant that scales time t, and H is the *Hurst exponent* (Smith, 2011). For a process to be self-similar, the value of H must fall in the range $0 \leq H \leq 1$. For a self-similar process is also to be long-range dependent, the value of H is restricted to $0.5 \leq H \leq 1$. A means of computing the Hurst parameter is given below; for a more detailed treatment of the subject, see (Clegg 2006).

One approach to demonstrating the existence of self-similar relationships in data is to produce a log-log plot, through analysis of data distributions from simulations (Fukada, Takayasu, and Takayasu, 2000; Fukada, Takayasu, and Takayasu, 2001). There are other techniques of course, and below, this appendix discusses various means for detecting and measuring self-similarity and long-range dependence, which were used by researchers surveyed in this study. The difficulties of using some of these techniques to obtain measurements are discussed in (Di Cairano-Gilfedder and Clegg, 2005; Clegg 2006).

Methods of Testing for Self-Similarity

First-Order Self-Similarity (Autocorrelation). In (Smith, 2011; Clegg, Di Cairano-Gilfedder, and Zhou, 2010), the well-known method for measuring first-order self-similarity is described. The procedure involves applying the autocorrelation function to a discrete time series for some stationary stochastic process, whose value X is measured at a series of discrete time step values, t, such that X = (Xt : t = 0, 1, 2, ... N), obtained either through periodic sampling or by averaging its value across a series of fixed length intervals:

$$p(s) = \frac{E\left[(X_t - \mu)(X_{t-s} - \mu)\right]}{\sigma^2} \qquad \text{(A2)}$$

where μ is the mean value of X and σ^2 is the variance. The variable s defines the *time lag*, $t - s$, which represents the number of time steps over which the autocorrelation coefficient, $p(s)$, is computed. If the process X is self-similar, $p(s)$, exhibits a power-law behavior

$$p(s) \sim ds^{-\beta}, \qquad \text{(A3)}$$

where $0 < \beta < 1$ and d is a constant. Equation (A3) approximates the asymptotic behavior of the system as s goes to infinity (or $s \rightarrow \infty$). Typically, the value of β can be estimated from measurements made using Equation (A2), employing a numerical method for this purpose. From (A3), it is possible to compute the Hurst exponent, as $H = 1 - \beta / 2$ (Clegg 2006) to determine the existence of long-range dependence (Arrowsmith et al., 2004).

<u>*Second-Order Self-Similarity (Aggregated Variance)*</u>. Following (Smith, 2011; Leland et al., 1994), second-order self-similarity, or *aggregated variance analysis*, is determined by re-aggregating the original time series for a stochastic process $X = (X_t : t = 0, 1, 2, ... N)$, using different "windows" of m consecutive values, as for example, $t = 0, m, 2m.., N / m$. Values in each series of windows are averaged. Thus, in the re-aggregated time series, each of the m values, $X_s^{(m)}$, is given by:

$$X\text{–}_s^{(m)} = 1 / m(X_{sm-m+1} \; + ... + X_{sm}) \qquad \text{(A4)}$$

for $m = 1, 2, 3…$, where s is as defined above. The autocorrelation coefficients of the re-aggregated time series are then compared to those of the original time series to determine if the correlation level is preserved. The re-aggregated time series is considered *exactly self-similar* if the variance Var $(X^{(m)}) = \sigma^2 / m^{-\beta}$ and $p^{(m)}(s) = p(s)$, for $s \geq 0$ (Smith, 2011), where $p^{(m)}$ is the autocorrelation measure defined by Equation (A3) with the exponent β, and σ^2 is computed from the original time series. In (Leland et al., 1994), exact self-similarity exists if the re-aggregated "processes $X^{(m)}$ have the same correlation structure", as X as determined through application of the autocorrelation function, $p(s)$. "In other words, X is exactly self-similar if the aggregated processes $X^{(m)}$ are indistinguishable from X—at least with respect to their second-order statistical properties" (Leland et al., 1994). The re-aggregated time series is considered *asymptotically self-similar* if $p^{(m)}(s) \rightarrow p(s)$, as $s \rightarrow \infty$ (Smith, 2011; Leland et al., 1994). Aggregated variance analysis was used in (Lawniczak et al., 2007) and to test for self-similarity in (Leland et al., 1994; Veres et al., 2003).

<u>*R/S Measurements*</u>. In this method, a time series for a stochastic process $X = (X_t : t = 0, 1, 2, ... N)$ is divided into m blocks of equal length N / m and the values in each block are summed. Setting $n = N / m$, the range $R(n)$ is defined as the difference between the value of the largest block and the smallest block, while $S(n)$ is the standard deviation of the summed values of the blocks. The ratio $R(n) / S(n)$ should scale with n such that

$$E\left[\frac{R(n)}{S(n)}\right] \sim gn^H \qquad \text{(A5)}$$

where g is a constant and H is the Hurst exponent as defined above (Smith, 2011). In (Leland et al., 1994), Equation (A5) was computed over the blocks of the re-aggregated time series from a number of different "starting points" to obtain a collection of sample R/S values. The *log of this sample of R/S values* was then plotted against *log* (n). Its slope is determined and matched against the data. Once a match was obtained, the asymptotic slope of this plot was used to estimate H and to determine the existence of self-similarity. This method was found to be relatively robust against "changes of the marginal distribution" (Leland et al., 1994). A different approach was taken in (Woolf et al., 2002) to estimate H using Equation (A5).

<u>*The 1/f Signal and Power-Spectrum Analysis*</u>. In a well-known paper by (Bak, Tang and Wiesenfield, 1987), it is argued that dynamical systems can evolve towards a *self-organized critical state*, in which they exhibit spatial and temporal power-law scaling behavior. Systems that reach the critical state "naturally evolve into self-organized critical structures of states which are barely stable" (Bak, Tang, and

Wiesenfeld, 1987). The power frequency distribution, or power spectrum, of measurable quantities within systems that are barely stable is described by the expression,

$$S(f)=cf^{-\alpha}, \qquad \text{(A6)}$$

where *c* is a constant and $\alpha \approx 1$. A distribution, which follows Equation (A6) is said to have a self-similar or "fractal" structure, in which $\alpha \approx 1$. Hence, the 1/*f* signal is regarded as a form of self-similarity. While other authors have generally subscribed to this view of the 1/*f* signal (Vespignani and Zepperi, 1998), there has been controversy (Milotti 2002), and the nature and source of the 1/*f* signal remains a topic of active research.

The existence of the 1/*f* signal has been reported in many domains, including domains related to natural processes, as for example, estimating risk of extinction from population changes (Halley and Kunin, 1999), electroencephalograms of human brain activity (Allegrini et al., 2009), as well as domains associated with human-engineered processes, such as electronic devices (Hooge, 1994) and musical pitch fluctuations (Levitan, Chordia, and Menon, 2012). Since the presence of the 1/*f* signal was reported in communication networks (Csabai, 1994), existence of this phenomenon has been reported in packet lifetimes, queue lengths, and other variables (Yuan and Mills, 2002; Mukherjee and Manna, 2005; Fukuda, Takayasu, and Takayasu, 2005).

For example, Figure 8 obtained from (Mukherjee and Manna, 2005) shows the 1/*f* pattern for queue lengths. In Figure 8 and in a system described by Equation (A6), the small number of large values reflects that the system is subjected to relatively few large perturbations, while the large number of small values, which increase exponentially with diminishing size, indicates that most perturbations are small. Arguably, the 1/*f* pattern, which is sometimes described as 1/*f* "noise", is in fact not noise, but a signal which describes the system state. If a system undergoes a phase transition, the number of large perturbations can be expected to increase, and the overall signal to change. As pointed out in Section 7, it is notable that 1/*f* noise was observed in these and other works both near and away from the critical point. To date, there is no complete analysis of which measureable quantities within communication networks exhibit the 1/*f* signal, or, more importantly, how the 1/*f* signal behaves in relation to the critical point (i.e., is it present or absent below, at, and above a critical point?). For this reason, the study of the 1/*f* signal, and its relationship to phase transitions in communication networks, remains an important topic of future work (see Section 7)—as is the case with respect to the topic of the 1/*f* signal in general.

Methods for measuring the behavior of systems and testing for the 1/*f* signal are numerous, and an in-depth treatment of this subject is beyond the scope of this survey. At a minimum, scientists may collect and de-trend data, which is then plotted as a power spectrum on a log-log graph. If the curve appears to be 1/*f* like, α may be estimated using a linear fitting technique (Ward and Greenwood, 2007). Beyond this, some more common methods for analyzing time series data to determine whether or not the 1/*f* signal is present, include use of discrete Fourier transform techniques combined with linear regression (Heinzel, Rudiger, and Schilling, 2002; Mukherjee and Manna, 2005); maximum likelihood estimation (Pilgram and Kaplan, 1998), and wavelet analysis (Ninness, 1998). Each of these methods is based on the use of a sophisticated mathematical model for describing the 1/*f* signal, for which parameter values must be estimated from the data of the system being analyzed. The parameterized model is then used to compute the value of $S(f)$ and determine whether the 1/*f* exists or not. The question of which of these methods to use in the communication network domain is a difficult one, and there appears to be no single answer. A brief introduction to the topic of estimating the 1/*f* signal is provided in (Ward and Greenwood, 2007).

<u>*Other Methods*</u>. Other methods, which are omitted due to space limitations, have been used besides those described above to measure self-similarity. These include wavelet methods (Abry and Veitch, 1998; Crovella and Kowalczyk, 2003), where wavelet methods were generally used in (Erramilli et al., 2002) and interval distribution of level (IDL) set analysis (Takayasu, 1993) used in (Fukuda, Takayasu, and Takayasu, 2005).

Methods for Testing for Long-Range Dependence

In (Clegg, 2006), long-range dependence is described as follows: "In the time domain it manifests as a high degree of correlation between distantly separated data points. In the frequency domain it manifests as a significant level of power at frequencies near zero", which echoes the description provided in (Karagiannis, Molle, and Faloutsos, 2004) cited at the beginning of this appendix. In addition to the use of the Hurst parameter, an indicator of long-range dependence may be obtained from the autocorrelation function, *p* (*s*), defined in Equation (A3), if

$$\sum_{s}^{\infty} p(s) \rightarrow \infty \tag{A7}$$

is found to diverge, where s defines a series of time lag values as indicated above (Smith, 2011; Clegg, Di Cairano-Gilfedder, and Zhou, 2010). The condition defined in (A7) may be satisfied regardless of whether or not the process is found to be self-similar, especially if the Hurst parameter cannot be defined. Thus, under certain conditions, one can have long-range dependence without self-similarity (Smith, 2011).

If second-order self-similarity is computed, i.e., aggregated variance, then long-range dependence exists if the aggregated process converges slowly to 0 at a rate below 1 / *m*, where the aggregated process was found to have a variance Var ($X^{(m)}$) = σ^2 / $m^{-\beta}$ and to be independent and identically distributed (Karagiannis, Molle, and Faloutsos, 2004) (or *exactly self-similar*; see previous section). Aggregated variance analysis and other methods (cited above) were used in (Lawniczak et al., 2007) to estimate the Hurst parameter, from which long-range dependence was determined. The Hurst exponent was also used to infer long-range dependence in (Solé and Valverde, 2001). Power spectral analysis also was used to infer long-range dependence in (Mukherjee and Manna, 2005; Tadic, Rodgers, and Thurner, 2007) and others. In (Arrowsmith et al., 2004), autocorrelation was computed as:

$$p\,(s) = \frac{E(X_t X_{t+s}) - E(X_t)E(X_{t+s})}{\sqrt{V(X_t)}\sqrt{V(X_{t+s})}} \tag{A8}$$

where *V* is the variance and 0 < *β* < 1 as in Equation (A3). Here, long-range dependence may be expressed using the Hurst parameter if 0.5 ≤ *H* ≤ 1 (Arrowsmith et al., 2004).

Long-range dependence may also be tested for by first using one of the above methods for calculating self-similarity to estimate the Hurst parameter, *H* = 1 – *β* / 2, where *β* may be estimated, for instance, from Equation (A3). This is the primary way in which most researchers detect the presence of long-range dependence. For a survey of methods for estimating the Hurst parameter (some of which give inconsistent results under certain conditions), see (Karagiannis, Molle, and Faloutsos, 2004) or (Clegg, 2006). In (Yuan and Mills, 2006) wavelet analysis (Abry and Vietch, 1998; Crovella and Kowalczyk, 2003) is used to study long-range dependence.